Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Stochastic Processes and their Applications 113 (2004) 235–272
stochastic processes and their applications
www.elsevier.com/locate/spa

Stochastic averaging and asymptotic behavior of the stochastic Duffing–van der Pol equation

Peter H. Baxendale*

Department of Mathematics, University of Southern California, 3620 S. Vermont Avenue, Los Angeles, CA 90089-2532, USA

Received 17 July 2002; received in revised form 7 April 2003; accepted 12 May 2004

Available online 15 June 2004

Abstract

Consider the stochastic Duffing–van der Pol equation

$$\ddot{x} = -\omega^2 x - Ax^3 - Bx^2\dot{x} + \varepsilon^2\beta\dot{x} + \varepsilon\sigma x\dot{W}_t$$

with $A \geqslant 0$ and $B > 0$. If $\beta/2 + \sigma^2/8\omega^2 > 0$ then for small enough $\varepsilon > 0$ the system $(x, \dot{x})$ is positive recurrent in $\mathbf{R}^2 \setminus \{0\}$. Let $\tilde{\lambda}_\varepsilon$ denote the top Lyapunov exponent for the linearization of this equation along trajectories. The main result asserts that

$$\tilde{\lambda}_\varepsilon \sim \varepsilon^2 \tilde{\lambda} \text{ as } \varepsilon \to 0,$$

where $\tilde{\lambda}$ is the top Lyapunov exponent along trajectories for a stochastic differential equation obtained from the stochastic Duffing–van der Pol equation by stochastic averaging. In the course of proving this result, we develop results on stochastic averaging for stochastic flows, and on the behavior of Lyapunov exponents and invariant measures under stochastic averaging. Using the rotational symmetry of the stochastically averaged system, we develop theoretical and numerical methods for the evaluation of $\tilde{\lambda}$. We see that the sign of $\tilde{\lambda}$, and hence the asymptotic behavior of the stochastic Duffing–van der Pol equation, depends strongly on $\omega B/A$. This dimensionless quantity measures the relative strengths of the nonlinear dissipation $Bx^2\dot{x}$ and the nonlinear restoring force Ax^3.

MSC: Primary 37H15; 60H10; 60J60; Secondary 37H20; 60F05; 60H35

Keywords: Stochastic differential equation; Stochastic averaging; Lyapunov exponents; Invariant measures; Stochastic flows; Duffing–van der Pol equation

* Tel.: +213-7403789; fax: +213-7402424.
E-mail address: baxendal@math.usc.edu (P.H. Baxendale).

0304-4149/$ - see front matter
doi:10.1016/j.spa.2004.05.001

1. Introduction

The noisy Duffing–van der Pol equation has recently been the object of much study in the theory of stability and bifurcations for stochastic dynamical systems. See for example Arnold [1], Arnold et al. [5], Keller and Ochs [14], and Schenk-Hoppé [26]. In this paper we will restrict attention to the system with multiplicative white noise. To be specific we consider the stochastic Duffing–van der Pol equation

$$\ddot{x} = -\omega^2 x + \beta\dot{x} - Ax^3 - Bx^2\dot{x} + \sigma x\dot{W}_t$$

with $\omega > 0$ and $A \geqslant 0$ and $B > 0$. Here $-\beta$ represents friction and σ is the intensity of the white noise forcing. Putting $y = -\dot{x}/\omega$ we get the 2-dimensional stochastic differential equation

$$\begin{aligned} \mathrm{d}x_t &= -\omega y_t\,\mathrm{d}t, \\ \mathrm{d}y_t &= (\omega x_t + \beta y_t + (A/\omega)x_t^3 - Bx_t^2 y_t)\,\mathrm{d}t - (\sigma/\omega)x_t\,\mathrm{d}W_t. \end{aligned} \tag{1}$$

Linearizing this system along the trajectory $\{(x_t, y_t)\colon t \geqslant 0\}$ in $\mathbf{R}^2$ gives the equation

$$\mathrm{d}v_t = \begin{bmatrix} 0 & -\omega \\ \omega + (3A/\omega)x_t^2 - 2Bx_t y_t & \beta - Bx_t^2 \end{bmatrix} v_t\,\mathrm{d}t - (\sigma/\omega)\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} v_t\,\mathrm{d}W_t. \tag{2}$$

Writing $v_t = \|v_t\| \begin{bmatrix} \cos\theta_t \\ \sin\theta_t \end{bmatrix}$ we obtain

$$\begin{aligned} \mathrm{d}\theta_t = {} & \left(\omega + \left(\frac{3A}{\omega}x_t^2 - 2Bx_t\,y_t\right)\cos^2\theta_t \right. \\ & \left. + (\beta - Bx_t^2)\sin\theta_t\cos\theta_t - \frac{\sigma^2}{\omega^2}\sin\theta_t\cos^3\theta_t\right)\mathrm{d}t \\ & - \frac{\sigma}{\omega}\cos^2\theta_t\,\mathrm{d}W_t. \end{aligned} \tag{3}$$

The (top) almost sure Lyapunov exponent $\lambda = \lambda(\beta, \sigma, \omega)$ for system (1) linearized at 0 is defined as the almost sure limit

$$\lambda(\beta, \sigma, \omega) = \lim_{t\to\infty} \frac{1}{t} \log \|v_t\|,$$

where $\{v_t\colon t \geqslant 0\}$ is a solution of (2) with $x_t = y_t = 0$ for all $t \geqslant 0$. It is easy to show that the limit exists almost surely and does not depend on v_0 so long as $v_0 \neq 0$. An exact formula for $\lambda(\beta, \sigma, \omega)$ is given by Imkeller and Lederer [13]; earlier numerical calculations by Kozin and Prodromou [19] gave criteria in terms of β, σ and ω for $\lambda(\beta, \sigma, \omega)$ to be positive or negative.

If $\lambda(\beta, \sigma, \omega) < 0$ then all solutions $\{(x_t, y_t)\colon t \geqslant 0\}$ of (1) converge to 0 almost surely as $t \to \infty$. Thus the fixed point 0 is almost surely stable. However if $\lambda(\beta, \sigma, \omega) > 0$ then the solution $\{(x_t, y_t)\colon t \geqslant 0\}$ of (1) is positive recurrent on $\mathbf{R}^2 \setminus \{0\}$ with a unique invariant probability measure $\mu(\beta, \sigma, \omega, A, B)$, say. (See Proposition 5.2 and the remark thereafter.) In this case we would like to describe the measure $\mu(\beta, \sigma, \omega, A, B)$.

Also we are interested in the question of stability along trajectories. If (x_0, y_0) and $(\tilde{x}_0, \tilde{y}_0)$ are distinct starting points in $\mathbf{R}^2/\{0\}$, and if $\{(x_t, y_t) : t \geqslant 0\}$ and $\{(\tilde{x}_t, \tilde{y}_t) : t \geqslant 0\}$ are the corresponding solutions of (1) generated by the same noise $\{W_t : t \geqslant 0\}$, then we are interested in the behavior of $\|(\tilde{x}_t, \tilde{y}_t) - (x_t, y_t)\|$ as $t \to \infty$. In this paper we shall consider the distinct but closely related question of linearized stability along trajectories. Instead of considering the behavior of $(\tilde{x}_t, \tilde{y}_t) - (x_t, y_t)$ we consider the behavior of the solution v_t of (2) when $\{(x_t, y_t) : t \geqslant 0\}$ is stationary in $\mathbf{R}^2/\{0\}$. In particular we consider the (top) almost sure Lyapunov exponent $\tilde{\lambda}(\beta, \sigma, \omega, A, B)$, say, for system (1) linearized along a stationary trajectory $\{(x_t, y_t) : t \geqslant 0\}$ in $\mathbf{R}^2/\{0\}$. The Lyapunov exponent $\tilde{\lambda}(\beta, \sigma, \omega, A, B)$ exists only when $\lambda(\beta, \sigma, \omega) > 0$, and then it is defined as the almost sure limit

$$\tilde{\lambda}(\beta, \sigma, \omega, A, B) = \lim_{t \to \infty} \frac{1}{t} \log \|v_t\|,$$

where $\{(x_t, y_t, v_t) : t \geqslant 0\}$ is a solution of (1), (2) with $\{(x_t, y_t) : t \geqslant 0\}$ stationary in $\mathbf{R}^2 \setminus \{0\}$.

In addition to the issue of stability along trajectories, the Lyapunov exponent $\tilde{\lambda}(\beta, \sigma, \omega, A, B)$ has significance for the (forward) stochastic flow generated by the stochastic differential equation (1). The sign of $\tilde{\lambda}(\beta, \sigma, \omega, A, B)$ should determine whether or not the random invariant measure for the (forward) stochastic flow generated by (1) is a random Dirac measure. This in turn has consequences for the random attractors associated with the stochastic flow. From the point of view of stochastic bifurcation theory it is of particular interest to know about the sign of $\tilde{\lambda}(\beta, \sigma, \omega, A, B)$ near a point in parameter space (β, σ, ω) where $\lambda(\beta, \sigma, \omega)$ changes sign. When the stable fixed point 0 becomes unstable, and a new (stable) random invariant measure appears on $\mathbf{R}^2 \setminus \{0\}$, it is important to know whether or not this new measure is atomic. For more details of these issues we refer the reader to Arnold [1, Chapter 9] and references therein.

The methods of Khasminskii [16] and Carverhill [11] (see Sections 4 and 5) give the formula

$$\tilde{\lambda}(\beta, \sigma, \omega, A, B) = \int_{(\mathbf{R}^2 \setminus \{0\}) \times \mathbf{S}^1} Q(x, y, \theta) \, \mathrm{d}P(x, y, \theta),$$

where

$$Q(x, y, \theta) = \left(\frac{3A}{\omega} x^2 - 2Bxy \right) \sin\theta \cos\theta + (\beta - Bx^2) \sin^2\theta + \frac{\sigma^2}{2\omega^2} \cos 2\theta \cos^2\theta$$

and P is the unique invariant probability measure for the diffusion $\{(x_t, y_t, \theta_t) : t \geqslant 0\}$ on $(\mathbf{R}^2 \setminus \{0\}) \times \mathbf{S}^1$ given by (1), (3). The existence of P is guaranteed by the existence of $\mu(\beta, \sigma, \omega, A, B)$ for $\lambda(\beta, \sigma, \omega) > 0$, and the uniqueness of P can be verified using the results of Arnold and San Martin [4]. To date there is no closed form formula for $\tilde{\lambda}(\beta, \sigma, \omega, A, B)$, and numerical evaluation is unreliable when attempting to distinguish very small positive values from very small negative values.

In this paper we adapt the stochastic averaging method used by Arnold et al. [5] to obtain rigorous asymptotic results for the invariant measure $\mu(\beta,\sigma,\omega,A,B)$ and the Lyapunov exponent $\tilde{\lambda}(\beta,\sigma,\omega,A,B)$ for small friction $-\beta$ and small noise intensity σ. We replace β by $\varepsilon^2\beta$ and σ by $\varepsilon\sigma$ to obtain

$$\begin{aligned} \mathrm{d}x_t &= -\omega y_t\,\mathrm{d}t \\ \mathrm{d}y_t &= \left(\omega x_t + \varepsilon^2\beta y_t + \frac{A}{\omega}x_t^3 - Bx_t^2 y_t\right)\mathrm{d}t - \frac{\varepsilon\sigma}{\omega}x_t\,\mathrm{d}W_t. \end{aligned} \tag{4}$$

In order to obtain a system which is a small perturbation of a rigid rotation we rescale the spatial variables by putting $\tilde{x}=x/\varepsilon$ and $\tilde{y}=y/\varepsilon$. Notice that this spatial rescaling will have no effect on the Lyapunov exponents, and involves a simple rescaling of invariant measures. Dropping the tildes, we get

$$\begin{aligned} \mathrm{d}x_t &= -\omega y_t\,\mathrm{d}t, \\ \mathrm{d}y_t &= \omega x_t\,\mathrm{d}t + \varepsilon^2\left(\beta y_t + \frac{A}{\omega}x_t^3 - Bx_t^2 y_t\right)\mathrm{d}t - \frac{\varepsilon\sigma}{\omega}x_t\,\mathrm{d}W_t. \end{aligned} \tag{5}$$

Let μ_ε denote the unique invariant measure for the one-point motion of (5), if it exists. Then $\mu(\varepsilon^2\beta,\varepsilon\sigma,\omega,A,B)(C)=\mu_\varepsilon(\varepsilon^{-1}C)$ for all Borel subsets C of $\mathbf{R}^2\setminus\{0\}$.

Using a stochastic averaging method, we associate with system (5) the 2-dimensional stochastic differential equation

$$\begin{aligned} \mathrm{d}z_t = {} & \left(\frac{\beta}{2} - \frac{B\|z_t\|^2}{8}\right) z_t\,\mathrm{d}t - \frac{3A\|z_t\|^2}{8\omega} Jz_t\,\mathrm{d}t \\ & + \frac{\sigma}{2\sqrt{2}\omega}K_1 z_t\,\mathrm{d}W_t^1 + \frac{\sigma}{2\sqrt{2}\omega}K_2 z_t\,\mathrm{d}W_t^2 + \frac{\sigma}{2\omega}Jz_t\,\mathrm{d}W_t^3, \end{aligned} \tag{6}$$

where

$$J=\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \quad \text{and} \quad K_1=\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \quad \text{and} \quad K_2=\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

System (6) is obtained by applying a rigid rotation through $-\omega t$ radians to the solutions of (5), rescaling time $t\mapsto t/\varepsilon^2$, and then letting $\varepsilon\to 0$. Details are given in Section 3. Since the rotation does not affect distances in $\mathbf{R}^2$ it is reasonable to believe that the asymptotic behavior of (6) as $t\to\infty$ will give information about the asymptotic behavior of (5) as $t\to\infty$ for small $\varepsilon>0$. In this paper we convert this act of faith into rigorous mathematical results for the invariant probability measure for the one-point motion on $\mathbf{R}^2\setminus\{(0,0)\}$, and for the almost sure Lyapunov exponent for the linearization along trajectories in $\mathbf{R}^2\setminus\{(0,0)\}$.

Let λ denote the almost sure Lyapunov exponent for the linearization of (6) at 0. If $\lambda>0$ let μ denote the invariant probability measure for the one-point motion of (6) on $\mathbf{R}^2\setminus\{(0,0)\}$, and let $\tilde{\lambda}$ denote the almost sure Lyapunov exponent for the linearization of (6) along trajectories in $\mathbf{R}^2\setminus\{(0,0)\}$. Our main result is the following.

Theorem 1.1. *Assume $A > 0$ and $B > 0$.*

(i) $\lambda(\varepsilon^2\beta, \varepsilon\sigma, \omega) \sim \varepsilon^2\lambda$ *as* $\varepsilon \to 0$, *where* $\lambda = \beta/2 + \sigma^2/8\omega^2$.
(ii) *If* $\lambda > 0$ *then* μ_ε *converges weakly to* μ *as* $\varepsilon \to 0$, *where with respect to 2-dimensional Lebesgue measure*

$$\mathrm{d}\mu(x,y) = cr^{8\omega^2\beta/\sigma^2}\exp\left(-\frac{B\omega^2 r^2}{\sigma^2}\right)\,\mathrm{d}x\,\mathrm{d}y.$$

Here $r = \sqrt{x^2+y^2}$ *and the normalizing constant* c *is given by*

$$c^{-1} = 4\pi\left(\frac{\sigma^2}{B\omega^2}\right)^{\omega^2\beta/2\sigma^2+1/8}\Gamma\left(\frac{\omega^2\beta}{2\sigma^2}+\frac{1}{8}\right).$$

(iii) *If* $\lambda > 0$ *then* $\tilde{\lambda}(\varepsilon^2\beta, \varepsilon\sigma, \omega, A, B) \sim \varepsilon^2\tilde{\lambda}$ *as* $\varepsilon \to 0$.

The estimate in (i) is an old result of Auslender and Milstein [6]. It implies that if $\lambda > 0$ then μ_ε and $\tilde{\lambda}(\varepsilon^2\beta, \varepsilon\sigma, \omega, A, B)$ exist for all sufficiently small $\varepsilon > 0$. The results (ii) and (iii) are applications of Theorems 5.1 and 5.2 to the stochastic differential equation (5). The verification that (5) satisfies the conditions of Theorems 5.1 and 5.2, and the derivations of (6) and the formula for μ are carried out in Section 6.

The advantage of using the new system (6) is that it has rotational symmetry. The symmetry is a direct consequence of the method used to obtain (6), and it implies that any calculation done for the system (6) will involve one dimension less than the corresponding calculation for (5). This is why we can give explicit formulas for λ and μ.

We do not have an exact formula for $\tilde{\lambda}$, since the symmetry of (6) enables us only to reduce a 3-dimensional integral expression to the 2-dimensional one given in (44). In Section 6 we use scaling arguments show that $\tilde{\lambda}$ is a function of σ/ω and β and $\omega B/A$ only. Therefore the effect of the nonlinear terms Ax^3 and $Bx^2\dot{x}$ in the original equation shows up only in the dimensionless quantity $\omega B/A$. The reduction in dimension from 3 to 2 makes numerical techniques much more feasible, and in Section 7 we report the results of Monte Carlo simulation using a first order Euler scheme to generate solutions of a 2-dimensional stochastic differential equation. The calculations indicate that in the range $-\sigma^2/4\omega^2 < \beta < 0$ we have

$$\tilde{\lambda} > 0 \quad \text{if } \omega B/A \leqslant 1,$$

$$\tilde{\lambda} < 0 \quad \text{if } \omega B/A \geqslant 1.5.$$

See Fig. 1. Thus if the coefficient B involved in the nonlinear dissipation term $x^2\dot{x}$ is large enough relative to the coefficient A involved in the nonlinear restoring term x^3 we have stability along trajectories for (1) for small enough $\varepsilon > 0$. Conversely if B is too small relative to A then we have instability along trajectories for (1) for small enough $\varepsilon > 0$.

The numerical calculations shown in Fig. 2 suggest the following bifurcation scenario for system (6) with $\omega B/A = 1.24$ when β is increased from $-\infty$ to 0. (Recall that $-\beta$ is the friction coefficient.) For $\beta < -\sigma^2/4\omega^2$ all trajectories converge to 0 almost surely, and so the fixed point is almost surely stable. For $-\sigma^2/4\omega^2 < \beta < -\beta_0\sigma^2/\omega^2$ the fixed point 0 is almost surely unstable and we have stability along positive recurrent trajectories in $\mathbf{R}^2 \setminus \{0\}$. For $-\beta_0\sigma^2/\omega^2 < \beta < 0$ the fixed point 0 remains almost surely unstable and we have instability along the positive recurrent trajectories in $\mathbf{R}^2 \setminus \{0\}$. Here the constant β_0 depends on the ratio $\omega B/A = 1.24$. The numerical value of β_0 is approximately 0.17.

The results above use the fact that $B > 0$. Suppose instead that $B = 0$ and $A > 0$ and $\beta < 0$. Part (i) of Theorem 1.1 remains valid since it ignores the nonlinearities. However, as soon as $\lambda > 0$ the process z_t given by (6) almost surely goes to ∞ exponentially quickly and the invariant probability measure μ fails to exist, although for $\varepsilon > 0$ the invariant probability measures μ_ε exist. Moreover, when $B = 0$ the Lyapunov exponent $\tilde{\lambda}(\varepsilon^2\beta, \varepsilon\sigma, \omega, A, 0)$ is not of order ε^2 as $\varepsilon \to 0$. A recent result of Baxendale and Goukasian [9] shows that

$$\tilde{\lambda}(\varepsilon^2\beta, \varepsilon\sigma, \omega, A, 0) \sim \varepsilon^{2/3}\bar{\lambda}$$

for some $\bar{\lambda} > 0$. In [9] the system (4) is regarded not as a small perturbation of a rotation but as a small perturbation of the nonlinear Hamiltonian system corresponding to $H(x, y) = \omega(x^2 + y^2)/2 + Ax^4/4\omega$.

We have three major tasks in this paper. The first is to describe the stochastic averaging algorithm by which the stochastic differential equation (6) was obtained from (5). The second is to obtain rigorous results relating the asymptotic and equilibrium behaviors of (5) and (6). The final task is to describe the asymptotic and equilibrium behavior of (6).

The first and second tasks will be carried out in the more general setting described at the beginning of Section 2. This setting includes not only our system (5) but also systems perturbed by small additive white noise. The passage from the general version of (5) to the general version of (6) is carried out in Section 3. It is important to be aware that the stochastic averaging is carried out for the n-point motions of the stochastic differential equations. See Theorem 3.1. Stochastic averaging is most commonly done just for the one-point motion. The information obtained as a result of stochastic averaging for the one-point motion of (5) is the generator for the one-point motion $\{z_t : t \geqslant 0\}$, and it is well known that there are many different choices of a stochastic differential equation which generate the same one-point motion. These choices will all be equally useful when we look for a result such as Theorem 1.1(ii) which involves just one-point motions. But these choices may produce very different limiting behavior for the 2-point motion, and for the motion linearized along trajectories. In particular these different choices may produce many different values of $\tilde{\lambda}$. However when stochastic averaging is done for the n-point motions the resulting information uniquely determines the local characteristics (see Kunita [20]) for the stochastic differential equation (6). In particular the law of the linearized process of (6) is uniquely determined, and so the Lyapunov exponent along trajectories for (6) is uniquely determined.

In Sections 4 and 5 we obtain results on the convergence of invariant measures and Lyapunov exponents. The results on convergence of Lyapunov exponents are proved via convergence of the measures in the Khasminskii–Carverhill formula. The majority of existing results on stochastic averaging deal with convergence on fixed time intervals $0 \leqslant t \leqslant T$ as $\varepsilon \to 0$. Thus they cannot be applied directly to questions involving invariant measures and behavior as $t \to \infty$. In this paper we take early stochastic averaging results of Khasminskii [15], including results about convergence of invariant measures, and adapt them to our setting. The most important extension involves the replacement of boundedness conditions on the coefficients by an assumption of the existence of suitable Lyapunov functions. These results appear in a self-contained appendix (Appendix A). The result (Theorem A.3) on convergence of invariant measures is applied in various slightly different settings to obtain Theorems 4.1, 4.2, 5.1 and 5.2. Section 4 deals with stochastic differential equations where 0 is not fixed, and then Section 5 adapts the results of Section 4 to the case when 0 is fixed. Thus Section 4 is relevant for systems with additive noise, and Section 5 is relevant for systems with multiplicative noise.

In Section 6 we return to the specific case of the stochastic Duffing–van der Pol equation with multiplicative noise. We derive (6) in the manner described in Section 3 and we verify that the assumptions of Section 2 are satisfied. We obtain the formula for the density of the invariant measure μ, and we discuss the Lyapunov exponent $\tilde{\lambda}$. Numerical results for $\tilde{\lambda}$ are given in Section 7.

2. Small perturbations of a rigid rotation

Much of the material in the paper applies to a wide class of systems dealing with small perturbations of a rigid rotation. From here until the end of Section 5 we will consider the system

$$dx_t^\varepsilon = -\omega J x_t^\varepsilon \, dt + \varepsilon^2 V_0(x_t^\varepsilon)\, dt + \varepsilon \sum_{\alpha=1}^{k} V_\alpha(x_t^\varepsilon)\, dW_t^\alpha, \tag{7}$$

where now $x_t^\varepsilon \in \mathbf{R}^2$ and $V_0, V_1, \ldots V_k$ are vector fields on $\mathbf{R}^2$. We shall assume throughout Sections 3 through 5 that the vector fields are C^3, although the interested reader will be able to identify places where C^2 or C^1 is sufficient.

Let L_ε denote the generator of the one-point motion associated with (7). Then

$$L_\varepsilon f(x) = -\omega Df(x)(Jx) + \varepsilon^2 Lf(x),$$

where

$$Lf(x) = \frac{1}{2}\sum_{\alpha=1}^{k} D^2 f(x)(V_\alpha(x), V_\alpha(x)) + Df(x)(V_0(x)).$$

The unperturbed system $\dot{x} = -\omega Jx$ has solution $x_t = R_{\omega t} x_0$ where $R_s = \left[\begin{smallmatrix} \cos s & -\sin s \\ \sin s & \cos s \end{smallmatrix}\right]$ denotes rigid rotation through s radians. Define the "averaged" operator $\bar{L}$ by

$$\bar{L} f(x) = \frac{1}{2\pi}\int_0^{2\pi} L(f \circ R_{-s})(R_s x)\, ds. \tag{8}$$

The definition of $\bar{L}$ implies that it is rotation invariant in the sense that

$$\bar{L}(f \circ R_{-s})(R_s x) = \bar{L} f(x).$$

In particular if f is a function of $\|x\|$ only, then so is $\bar{L} f$.

At various stages in Sections 3 through 5 we will assume one or more of the following hypotheses.

A1. $V_\alpha(0)=0$ for $0\leqslant\alpha\leqslant k$ and there exists $\alpha\geqslant 1$ and $u\in\mathbf{R}^2$ such that $\langle u, DV_\alpha(0)u\rangle \neq 0$.

A2. There exists $c<\infty$ and $\varepsilon_0>0$ and C^2 functions $F_\varepsilon(x)$ for $0<\varepsilon\leqslant\varepsilon_0$ satisfying $\inf_{0<\varepsilon\leqslant\varepsilon_0} F_\varepsilon(x)\to\infty$ as $\|x\|\to\infty$ and $\sup_{0<\varepsilon\leqslant\varepsilon_0}\sup_{x\in K} F_\varepsilon(x)<\infty$ for any compact $K\subset\mathbf{R}^2$ and

$$L_\varepsilon F_\varepsilon(x)\leqslant \varepsilon^2 c F_\varepsilon(x) \quad \text{for all } 0<\varepsilon\leqslant\varepsilon_0.$$

A3. There exists $c<\infty$ and a C^2 function $F_0(x)$ satisfying $F_0(x)\to\infty$ as $\|x\|\to\infty$ and

$$\bar{L}F_0(x)\leqslant cF_0(x).$$

A4. There exists a continuous function $G(x)\geqslant 0$ and constants $c<\infty$ and $0<R_1<R_2<\infty$ and $\varepsilon_0>0$ with the following properties:
(i) $G(x)\to\infty$ as $\|x\|\to\infty$,
(ii) for each $0<\varepsilon\leqslant\varepsilon_0$ there exists a non-negative C^2 function $F_\varepsilon(x)$ such that

$$\sup_{R_1\leqslant\|x\|\leqslant R_2}\left(|F_\varepsilon(x)|+\sum_{\alpha=1}^{k}|(V_\alpha.F_\varepsilon)(x)|\right)\leqslant c$$

and

$$L_\varepsilon F_\varepsilon(x)\leqslant -\varepsilon^2 G(x) \text{ whenever } \|x\|\geqslant R_1,$$

Remarks. (i) In A4 the notation $V_\alpha.F_\varepsilon$ denotes the vector field V_α acting as a first order differential operator on the function F_ε. At the cost of replacing $G(x)$ by $\min\{G(y): \|y\|=\|x\|\}$ we can and will assume that $G(x)$ is a function of $\|x\|$ only.

(ii) The assumption A3 ensures that the diffusion with generator $\bar{L}$ does not explode, see Khasminskii [17, Theorem III.4.1]. The assumption A2 is a "uniform non-explosion" condition. The assumption A4 is a "uniform positive recurrence" condition, see Khasminskii [17, Theorems III.7.3 and IV.4..1].

(iii) Except for part of hypothesis A1, we have not included non-degeneracy conditions in our list of assumptions. As a consequence we will have to include assumptions about uniqueness of certain invariant measures as part of the statements of our results. Easily verifiable conditions for the uniqueness of invariant probability measures are given in Arnold and Kliemann [3].

3. Stochastic averaging and flows

In this section we describe the passage from (7), the generalized version of (5), to a new stochastic differential equation (9), the generalized version of (6).

Define $a(x,y)\in\mathbf{R}^2\otimes\mathbf{R}^2$ and $U_0(x)\in\mathbf{R}^2$ by averaging as follows

$$a(x,y)=M_s\left\{\sum_{\alpha=1}^{k}R_{-s}V_\alpha(R_sx)\otimes R_{-s}V_\alpha(R_sy)\right\},$$

$$U_0(x)=M_s\{R_{-s}V_0(R_sx)\},$$

where M_s is the averaging operator $M_s\{f(s)\}=(1/2\pi)\int_0^{2\pi}f(s)\,\mathrm{d}s$. Notice that the operator $\bar{L}$ defined in (8) is given by

$$\bar{L}f(x)=\frac{1}{2}D^2f(x)(a(x,x))+Df(x)(U_0(x)).$$

The function a is a non-negative definite kernel with values in $\mathbf{R}^2\otimes\mathbf{R}^2$, and so it is the reproducing kernel for some separable Hilbert space $\mathscr{H}$ consisting of vector fields on $\mathbf{R}^2$. Moreover the function $a(x,y)$ is C^3 in each of its variables. Therefore $\mathscr{H}$ consists of C^3 vector fields. If U_β, $\beta\geqslant 1$, is an orthonormal basis for this Hilbert space, then we can write

$$a(x,y)=\sum_{\beta\geqslant 1}U_\beta(x)\otimes U_\beta(y).$$

Now consider the stochastic differential equation in $\mathbf{R}^2$

$$\mathrm{d}x_t=U_0(x_t)\,\mathrm{d}t+\sum_{\beta\geqslant 1}U_\beta(x_t)\,\mathrm{d}W_t^\beta, \tag{9}$$

where the $\{W_t^\beta : t\geqslant 0\}$ are independent standard Brownian motion processes. If the basis U_β is infinite this equation should be interpreted in the sense of Kunita [20], using the vector space valued Wiener process $Z_t=\sum_{q\geqslant 1}U_\beta W_t^\beta$. In this case the local characteristic (or covariance function) associated with Z_t is exactly $a(x,y)$. Notice however that if all the vector fields $V_\alpha(x)$ for $1\leqslant\alpha\leqslant k$ have coefficients which are polynomial functions of the coordinates of x of degree at most N, then the same will be true of all the vector fields in the reproducing kernel Hilbert space $\mathscr{H}$. Thus in this case $\mathscr{H}$ will be finite dimensional and the basis U_β will be finite.

For future reference we list the following formulas relating the coefficients of the original system (7) and the new system (9). Eqs. (10) and (12) are essentially restatements of the definitions of U_0 and the U_β, $\beta\geqslant 1$. The remaining equations are then obtained by differentiation of (10) and (12).

$$\langle U_0(x),u\rangle=M_s\{\langle V_0(R_sx),R_su\rangle\}, \tag{10}$$

$$\langle DU_0(x)\tilde{u},u\rangle=M_s\{\langle DV_0(R_sx)R_s\tilde{u},R_su\rangle\}, \tag{11}$$

$$\sum_{\beta\geqslant 1}\langle U_\beta(x),u\rangle\langle U_\beta(y),v\rangle=M_s\left\{\sum_{\alpha=1}^{k}\langle V_\alpha(R_sx),R_su\rangle\langle V_\alpha(R_sy),R_sv\rangle\right\}, \tag{12}$$

$$\sum_{\beta\geqslant 1}\langle DU_\beta(x)\tilde{u},u\rangle\langle U_\beta(y),v\rangle=M_s\left\{\sum_{\alpha=1}^{k}\langle DV_\alpha(R_sx)R_s\tilde{u},R_su\rangle\langle V_\alpha(R_sy),R_sv\rangle\right\}, \tag{13}$$

$$\sum_{\beta\geqslant 1}\langle DU_\beta(x)\tilde{u},u\rangle\langle DU_\beta(y)\tilde{v},v\rangle$$
$$= M_s\left\{\sum_{\alpha=1}^{k}\langle DV_\alpha(R_sx)R_s\tilde{u},R_su\rangle\langle DV_\alpha(R_sy)R_s\tilde{v},R_sv\rangle\right\}. \tag{14}$$

It follows that each of the terms on the left sides above are unchanged if all of the variables $x,y,u,v,\tilde{u},\tilde{v}$ are simultaneously rotated through the same angle.

Our construction of the stochastic differential equation (9) has been based on the observation that the stochastic differential equation (7) consists of a fast motion, the rotation $R_{\omega t}$, and that everything else is moving slowly relative to this rotation. Let ξ_t^ε denote the (local) stochastic flow associated with (7), and define $\tilde{\xi}_t^\varepsilon = R_{-\omega t}\xi_t^\varepsilon$. Then $\tilde{\xi}_t^\varepsilon$ is the flow along the time dependent SDE

$$\mathrm{d}\tilde{x}_t^\varepsilon = \varepsilon^2 R_{-\omega t}V_0(R_{\omega t}\tilde{x}_t^\varepsilon)\,\mathrm{d}t + \varepsilon\sum_{\alpha=1}^{k}R_{-\omega t}V_\alpha(R_{\omega t}\tilde{x}_t^\varepsilon)\,\mathrm{d}W_t^\alpha. \tag{15}$$

Theorem 3.1. *Assume* A2 *and* A3. *For any* $n\geqslant 1$, *let the* $\mathbf{R}^{2n}$*-valued process* $\{(\tilde{x}_t^{\varepsilon,1},\tilde{x}_t^{\varepsilon,2},\dots,\tilde{x}_t^{\varepsilon,n})\colon t\geqslant 0\}$ *denote the n-point motion determined by Eq.* (15). *For any* $T<\infty$ *the time-rescaled process* $\{(\tilde{x}_{t/\varepsilon^2}^{\varepsilon,1},\tilde{x}_{t/\varepsilon^2}^{\varepsilon,2},\dots,\tilde{x}_{t/\varepsilon^2}^{\varepsilon,n}) : 0\leqslant t\leqslant T\}$ *converges weakly in* $C([0,T],\mathbf{R}^{2n})$ *to the n-point motion of Eq.* (9).

Proof. We generalize the construction of (15) as follows. Define a process $\varphi_t\in\mathbf{S}^1=\mathbf{R}/2\pi\mathbf{Z}$ by

$$\mathrm{d}\varphi_t = \omega\,\mathrm{d}t \tag{16}$$

and then consider $\tilde{\xi}_t^\varepsilon = R_{-\varphi_t}\xi_t^\varepsilon$. This leads to the system

$$\mathrm{d}\tilde{x}_t^\varepsilon = \varepsilon^2 R_{-\varphi_t}V_0(R_{\varphi_t}\tilde{x}_t^\varepsilon)\,\mathrm{d}t + \varepsilon\sum_{\alpha=1}^{k}R_{-\varphi_t}V_\alpha(R_{\varphi_t}\tilde{x}_t^\varepsilon)\,\mathrm{d}W_t^\alpha. \tag{17}$$

Notice that (15) is the special case of (17) corresponding to $\varphi_0=0$. Henceforth in this proof for ease of notation we write $\tilde{x}_t^\varepsilon = y_t$. Let $y^{(1)},\dots,y^{(n)}$ be n points in $\mathbf{R}^2$ and let $\vec{y}$ be the corresponding vector in $\mathbf{R}^{2n}$. The equation for the n-point motion associated to (17) can be written

$$\mathrm{d}\vec{y}_t = \varepsilon^2\vec{V}_0(\varphi_t,\vec{y}_t)\,\mathrm{d}t + \varepsilon\sum_{\alpha=1}^{k}\vec{V}_\alpha(\varphi_t,\vec{y}_t)\,\mathrm{d}W_t^\alpha, \tag{18}$$

where the $(2i-1,2i)$ entries of $\vec{V}_\alpha(\varphi,\vec{y})$ are $R_{-\varphi}V_\alpha(R_\varphi y^{(i)})$ for $0\leqslant\alpha\leqslant k$ and $1\leqslant i\leqslant n$. We now apply Theorem A.2 from the appendix to the process $(\varphi_t,\vec{y}_t)\in\mathbf{S}^1\times\mathbf{R}^{2n}$ given by (16) and (18). The $(2i-1,2i)$ block of the drift vector for (18) is $R_{-\varphi}V_0(R_\varphi y^{(i)})$. When this is averaged over φ we obtain $U_0(y^{(i)})$, which is the $(2i-1,2i)$ block of the drift vector for the n-point motion of (9). The $(2i-1,2i)\times(2j-1,2j)$ block of the diffusion matrix corresponding to (18) is

$$\sum_{\alpha=1}^{k} R_{-\varphi} V_\alpha(R_\varphi y^{(i)}) \otimes R_{-\varphi} V_\alpha(R_\varphi y^{(j)}).$$

When this is averaged over φ we obtain $a(y^{(i)}, y^{(j)}) = \sum_\beta U_\beta(y^{(i)}) \otimes U_\beta(y^{(j)})$, which is the $(2i-1, 2i) \times (2j-1, 2j)$ block of the diffusion matrix for the n-point motion of (9). Therefore the operator which corresponds to the $\bar{L}$ appearing in Theorem A.2 is exactly the generator for the n-point motion of (9). It remains only to verify the conditions B2 and B3 of Theorem A.2. The condition B2 can be checked by direct calculation using the function $V_\varepsilon(\varphi, \vec{y}) = \sum_{i=1}^n F_\varepsilon(R_\varphi y^{(i)})$ where F_ε is the function which appears in assumption A2. Similarly, the condition B3 can be checked using the function $V_0(\vec{y}) = \sum_{i=1}^n F_0(y^{(i)})$ where F_0 is the function which appears in A3. □

With stronger boundedness assumptions on the coefficients, Theorem 3.1 would become a special case of Kunita's results on the convergence of stochastic flows under stochastic averaging, see [20] Section 5.6. Notice that Kunita has results on "convergence as diffusions" dealing with the n-point motions for all $n \geqslant 1$, as well as results on "convergence as flows". The results on convergence as flow would need stronger assumptions on the coefficients. In particular it would be necessary to ensure that Eqs. (15) and (9) have stochastic flows which exist for all time. Our result is most closely related to Theorem 5.5.1 of [20], but we have replaced boundedness assumptions on the coefficients by the assumptions A2 and A3.

Notice that stochastic averaging applied to the one point motion $t \to \tilde{\xi}^\varepsilon_{t/\varepsilon^2}(x)$ determines the vector field $U_0(x)$ and the values $a(x,x)$, and hence it determines the generator $\bar{L}$, but it does not determine the values $a(x,y)$ for $x \neq y$. It is easy to check that $\bar{L}$ is the generator of the one-point motion of (9), but $\bar{L}$ is insufficient to determine the law of the n-point motion of (9). When 0 is a fixed point, $\bar{L}$ does determine the law of the process $\{u_t : t \geqslant 0\}$ in $\mathbf{R}^2$ obtained by linearizing (9) at 0. In the presence of enough noise, this is enough to determine the top (almost sure) Lyapunov exponent and the moment Lyapunov exponent. It is not enough to determine the remainder of the Lyapunov spectrum for the linearization of (9) at 0. More significantly $\bar{L}$ is insufficient to determine the law of the process obtained by linearizing (9) along non-trivial stationary trajectories. In particular it is insufficient to determine the top Lyapunov exponent for the linearization of (9) along trajectories.

Theorem 3.1 deals with the behavior of solutions of (7) and (9) over finite intervals of time. In the next two sections we study the equilibrium behavior and the behavior as $t \to \infty$ of solutions of these two systems.

4. Additive noise

In Sections 2 and 3 we have made no assumptions as to whether or not the vector fields V_α fix 0. Suppose now that $V_\alpha(0) \neq 0$ for some $\alpha \geqslant 1$. Clearly 0 is not a fixed point for system (7). Since

$$a(0,0) = \frac{1}{2}\sum_{\alpha=1}^{k} \|V_\alpha(0)\|^2 \left(\begin{bmatrix} 1 \\ 0 \end{bmatrix} \otimes \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} \otimes \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right),$$

we deduce that 0 is not a fixed point for the averaged system (9). Typically in these circumstances there is at most one invariant probability measure μ for the one point motion of system (9). This will be the setting for the results in this section. Therefore this section is relevant for systems perturbed by additive noise. In particular it will apply to the case of the stochastic Duffing–van der Pol equation with additive noise.

In Section 5 we will consider the case when $V_\alpha(0)=0$ for all $\alpha \geqslant 0$. Although this is the case which includes our system (5), the fact that we need to work on $\mathbf{R}^2 \setminus \{0\}$ rather than $\mathbf{R}^2$ causes some extra technical complications. Thus we have chosen to postpone this case until after we have presented the main ideas in Section 4. Notice that we will not consider the case when $V_0(0) \neq 0$ and $V_\alpha(0)=0$ for all $\alpha \geqslant 1$. In this case 0 is fixed for the averaged system (9) but not for the original system (7).

We consider first invariant measures for the one point motions of (7) and (9). The one point motions of (7) and (9) are diffusions in $\mathbf{R}^2$ with generators L_ε and $\bar{L}$ respectively.

Lemma 4.1. *Assume* A2 *and* A4. *There is* $\varepsilon_0 > 0$ *such that for* $0 < \varepsilon \leqslant \varepsilon_0$ *there exists an invariant probability measure* μ_ε *for the diffusion with generator* L_ε. *Moreover there exists* $c_1 < \infty$ *such that any invariant probability measure* μ_ε *for* $0 < \varepsilon \leqslant \varepsilon_0$ *satisfies*

$$\int G(x)\,\mathrm{d}\mu_\varepsilon(x) \leqslant c_1.$$

Proof. Note first that assumption A2 implies that the diffusion with generator L_ε is non-explosive. Let ε_0 and F_ε and $0 < R_1 < R_2$ be as in assumption A4. Henceforth in this proof we take $0 < \varepsilon \leqslant \varepsilon_0$. Let $h(x)$ be a smooth radially symmetric function $\mathbf{R}^2 \to [0,1]$ with the properties that $h(x)=0$ if $\|x\| \leqslant R_1$ and $h(x)=1$ if $\|x\| \geqslant R_2$. Define the function $\tilde{F}_\varepsilon(x) = h(x)F_\varepsilon(x)$. Then $\tilde{F}_\varepsilon(x) \geqslant 0$ and

$$L_\varepsilon \tilde{F}_\varepsilon(x) = h(x)L_\varepsilon F_\varepsilon(x) + \varepsilon^2 Lh(x)F_\varepsilon(x) + \varepsilon^2 \sum_{\alpha=1}^{k} (V_\alpha.h)(x)(V_\alpha.F_\varepsilon)(x).$$

We can estimate the right side above using the facts that $h(x)=0$ when $\|x\| \leqslant R_1$ and that $Lh(x)=0$ and $(V_\alpha.h)(x)=0$ except when $R_1 \leqslant \|x\| \leqslant R_2$. Using the assumption A4 we get

$$L_\varepsilon \tilde{F}_\varepsilon(x) \leqslant -\varepsilon^2 h(x)G(x) + \varepsilon^2 c_1$$

for some constant $c_1 < \infty$. It follows from Meyn and Tweedie [21, Theorem 4.5] that the diffusion with generator L_ε has an invariant measure μ_ε, and that any invariant measure μ_ε satisfies

$$\int h(x)G(x)\,\mathrm{d}\mu_\varepsilon(x) \leqslant c_1.$$

Therefore

$$\int G(x)\,\mathrm{d}\mu_\varepsilon(x) \leqslant c_1 + \sup\{G(x) : \|x\| \leqslant R_2\}$$

and we are done. □

Theorem 4.1. *Assume* A2, A3 *and* A4. *Suppose the diffusion with generator $\bar{L}$ has at most one invariant probability measure. Then the diffusion with generator $\bar{L}$ has a unique invariant probability measure μ, say, and μ_ε converges weakly to μ as $\varepsilon \to 0$. Moreover, if $f(x)$ is any continuous function such that $f(x)/G(x) \to 0$ as $\|x\| \to \infty$ then $f \in L^1(\mu)$ and*

$$\int f(x)\,\mathrm{d}\mu_\varepsilon(x) \to \int f(x)\,\mathrm{d}\mu(x)$$

as $\varepsilon \to 0$.

Proof. We restrict to $0 < \varepsilon \leqslant \varepsilon_0$ where ε_0 is as in Lemma 4.1, and let μ_ε be an invariant probability measure for L_ε. Define the probability measure $\tilde{\mu}_\varepsilon$ on $\mathbf{S}^1 \times \mathbf{R}^2$ by

$$\int f(\varphi, x)\,\mathrm{d}\tilde{\mu}_\varepsilon(\varphi, x) = \int f(\varphi, R_{-\varphi}x)\,\mathrm{d}\mu_\varepsilon(x)\,\mathrm{d}m(\varphi),$$

where m denotes the uniform probability measure (Haar measure) on $\mathbf{S}^1$. Equivalently, $\tilde{\mu}_\varepsilon$ is the joint law of $(\varphi, R_{-\varphi}x)$ when φ and x are chosen independently with distributions m and μ_ε respectively. It follows directly that $\tilde{\mu}_\varepsilon$ is an invariant probability for the process $(\varphi_t, \tilde{x}_t^\varepsilon) \in \mathbf{S}^1 \times \mathbf{R}^2$ given by (16), (17) in the proof of Theorem 3.1. Assuming without loss of generality that the function $G(x)$ in assumption A4 is radially symmetric, Lemma 4.1 gives

$$\int_{\mathbf{S}^1 \times \mathbf{R}^2} G(x)\,\mathrm{d}\tilde{\mu}_\varepsilon(\varphi, x) = \int_{\mathbf{R}^2} G(x)\,\mathrm{d}\mu_\varepsilon \leqslant c_1.$$

This gives tightness of the $\tilde{\mu}_\varepsilon$. We now apply Theorem A.3 to obtain the existence of the unique μ for the diffusion with generator $\bar{L}$ and also the weak convergence of $\tilde{\mu}_\varepsilon$ to $m \times \mu$ as $\varepsilon \to 0$. Given bounded continuous $h : \mathbf{R}^2 \to \mathbf{R}$, define $f(\varphi, x) = h(R_\varphi x)$. The weak convergence above gives

$$\int h(x)\,\mathrm{d}\mu_\varepsilon(x) = \int f(\varphi, x)\,\mathrm{d}\tilde{\mu}_\varepsilon(\varphi, x)$$

$$\to \int \left(\int h(R_\varphi x)\,\mathrm{d}\mu(x) \right) \mathrm{d}m(\varphi) = \int h(x)\,\mathrm{d}\mu(x),$$

where the last equality uses the fact that μ is rotation invariant. This gives the weak convergence $\mu_\varepsilon \Rightarrow \mu$. The last statement is now a direct consequence of this weak convergence together with the fact that $\int G(x)\,\mathrm{d}\mu_\varepsilon$ is bounded as $\varepsilon \to 0$. □

Remark. The averaged system (9) has rotational symmetry. If the one-point motion of (9) has an invariant measure μ, then it must have a rotationally invariant measure.

So if μ is unique, it must be rotationally invariant. Thus μ can be obtained by solving a 1-dimensional problem.

We now consider the issue of stability along trajectories. If we linearize system (7) along a trajectory x_t^ε we get the linear stochastic differential equation

$$\mathrm{d}v_t^\varepsilon = -\omega J v_t^\varepsilon\,\mathrm{d}t + \varepsilon^2 DV_0(x_t^\varepsilon)v_t^\varepsilon\,\mathrm{d}t + \varepsilon\sum_{\alpha=1}^k DV_\alpha(x_t^\varepsilon)v_t^\varepsilon\,\mathrm{d}W_t^\alpha. \tag{19}$$

We write v_t^ε in polar coordinates $\|v_t^\varepsilon\|$ and $\theta_t^\varepsilon \in \mathbf{R}/2\pi\mathbf{Z}$. For any $\theta\in\mathbf{R}/2\pi\mathbf{Z}$ let $\bar\theta$ denote the corresponding unit vector $\bar\theta = \left[\begin{smallmatrix}\cos\theta\\ \sin\theta\end{smallmatrix}\right]$. Itô's formula gives

$$\begin{aligned}\mathrm{d}\theta_t^\varepsilon &= \omega\,\mathrm{d}t - \varepsilon^2\langle J\overline{\theta_t^\varepsilon}, DV_0(x_t^\varepsilon)\overline{\theta_t^\varepsilon}\rangle\,\mathrm{d}t\\ &\quad + \varepsilon^2\sum_{\alpha=1}^k\langle J\overline{\theta_t^\varepsilon}, DV_\alpha(x_t^\varepsilon)\overline{\theta_t^\varepsilon}\rangle\langle\overline{\theta_t^\varepsilon}, DV_\alpha(x_t^\varepsilon)\overline{\theta_t^\varepsilon}\rangle\,\mathrm{d}t\\ &\quad - \varepsilon\sum_{\alpha=1}^k\langle J\overline{\theta_t^\varepsilon}, DV_\alpha(x_t^\varepsilon)\overline{\theta_t^\varepsilon}\rangle\,\mathrm{d}W_t^\alpha\end{aligned} \tag{20}$$

and

$$\begin{aligned}\mathrm{d}\log\|v_t^\varepsilon\| &= \varepsilon^2\langle\overline{\theta_t^\varepsilon}, DV_0(x_t^\varepsilon)\overline{\theta_t^\varepsilon}\rangle\,\mathrm{d}t + \frac{\varepsilon^2}{2}\sum_{\alpha=1}^k\|DV_\alpha(x_t^\varepsilon)\overline{\theta_t^\varepsilon}e\|^2\,\mathrm{d}t\\ &\quad - \varepsilon^2\sum_{\alpha=1}^k\langle\overline{\theta_t^\varepsilon}, DV_\alpha(x_t^\varepsilon)\overline{\theta_t^\varepsilon}\rangle^2\,\mathrm{d}t + \varepsilon\sum_{\alpha=1}^k\langle\overline{\theta_t^\varepsilon}, DV_\alpha(x_t^\varepsilon)\overline{\theta_t^\varepsilon}\rangle\,\mathrm{d}W_t^\alpha.\end{aligned} \tag{21}$$

We see that the process $\{(x_t^\varepsilon,\theta_t^\varepsilon) : t\geqslant 0\}$ given by (7) and (20) is a diffusion process on $\mathbf{R}^2\times\mathbf{S}^1$. Under assumptions A2 and A4 the process $\{x_t^\varepsilon\colon t\geqslant 0\}$ on $\mathbf{R}^2$ has at least one invariant probability measure μ_ε. The compactness of $\mathbf{S}^1$ and the fact that $\{(x_t^\varepsilon,\theta_t^\varepsilon)\colon t\geqslant 0\}$ is a Feller process then implies that it has an invariant probability measure v_ε, say, on $\mathbf{R}^2\times\mathbf{S}^1$ with $\mathbf{R}^2$ marginal μ_ε.

Proposition 4.1. *Assume* A2 *and* A4. *Assume also that*

$$\limsup_{\|x\|\to\infty}\frac{\|DV_0(x)\| + \sum_{\alpha=1}^k\|DV_\alpha(x)\|^2}{G(x)} < \infty, \tag{22}$$

and that the process $\{(x_t^\varepsilon,\theta_t^\varepsilon) : t\geqslant 0\}$ on $\mathbf{R}^2\times\mathbf{S}^1$ given by (7) *and* (20) *has at most one invariant probability measure. Then the process given by* (7) *and* (20) *has a unique invariant probability measure v_ε, say. Let μ_ε denote the $\mathbf{R}^2$ marginal of v_ε. The almost sure Lyapunov exponent for system* (7)

$$\tilde\lambda_\varepsilon = \lim_{t\to\infty}\frac{1}{t}\log\|v_t^\varepsilon\| \tag{23}$$

exists with probability 1 *for* μ_ε*-almost all* x_0^ε *and all* $v_0^\varepsilon \neq 0$, *and is given by the formula*

$$\tilde{\lambda}_\varepsilon = \varepsilon^2 \int Q_V(x,\theta)\,\mathrm{d}v_\varepsilon(x,\theta), \tag{24}$$

where

$$Q_V(x,\theta) = \langle\bar{\theta}, DV_0(x)\bar{\theta}\rangle + \sum_{\alpha=1}^{k}\left(\frac{1}{2}\,\|DV_\alpha(x)\bar{\theta}\|^2 - \langle\bar{\theta}, DV_\alpha(x)\bar{\theta}\rangle^2\right).$$

Proof. As discussed above, Lemma 4.1 gives the existence of v_ε. Let Φ_t^ε denote the fundamental matrix solution for the linear stochastic differential equation (19). The assumption (22) together with Lemma 4.1 implies that $\|DV_0(x)\| \in L^1(\mu_\varepsilon)$ and $\|DV_\alpha(x)\|^2 \in L^1(\mu_\varepsilon)$ for all $\alpha \geqslant 1$. It follows (see Arnold [1] Theorem 4.2.13) that the integrability condition in the multiplicative ergodic theorem for $\{\Phi_t^\varepsilon : t \geqslant 0\}$ is satisfied. Therefore the limit in (23) exists with probability 1 for μ_ε-almost all x_0^ε, although at this stage in the argument the value may depend on v_0^ε. Eq. (21) gives

$$\log\|v_t^\varepsilon\| = \log\|v_0^\varepsilon\| + \varepsilon^2\int_0^t Q_V(x_s^\varepsilon, \theta_s^\varepsilon)\,\mathrm{d}s + M_t$$

for a continuous martingale M_t. Assumption (22) together with Lemma 4.1 gives v_ε-integrability conditions on Q_V and on the quadratic variation of M_t. So by the usual Khasminskii–Carverhill argument (see [16, 11]) we obtain (24) with probability 1 for v_ε-almost all $(x_0^\varepsilon, \theta_0^\varepsilon)$. Finally we use the uniqueness of v_ε. The argument used in Carverhill [11] Proposition 4.2 for the linearization about a fixed point applies also in our present setting for the linearization along trajectories. We deduce that formulas (23) and (24) are both valid with probability 1 for μ_ε-almost all x_0^ε and all $v_0^\varepsilon \neq 0$. □

Similarly we can linearize the averaged system (9) along a trajectory x_t to obtain

$$\mathrm{d}v_t = DU_0(x_t)v_t\,\mathrm{d}t + \sum_{\beta\geqslant 1} DU_\beta(x_t)v_t\,\mathrm{d}W_t^\beta. \tag{25}$$

Passing to polar coordinates $\|v_t\|$ and θ_t we get

$$\begin{aligned}\mathrm{d}\theta_t = {}& -\langle J\overline{\theta_t}, DU_0(x_t)\overline{\theta_t}\rangle\,\mathrm{d}t + \sum_{\beta\geqslant 1}\langle J\overline{\theta_t}, DU_\beta(x_t)\overline{\theta_t}\rangle\langle\overline{\theta_t}, DU_\beta(x_t)\overline{\theta_t}\rangle\,\mathrm{d}t\\ &-\sum_{\beta\geqslant 1}\langle J\overline{\theta_t}, DU_\beta(x_t)\overline{\theta_t}\rangle\,\mathrm{d}W_t^\beta\end{aligned} \tag{26}$$

and

$$\begin{aligned}\mathrm{d}\log\|v_t\| = {}& \langle\overline{\theta_t}, DU_0(x_t)\overline{\theta_t}\rangle\,\mathrm{d}t + \frac{1}{2}\sum_{\beta\geqslant 1}\|DU_\beta(x_t)\overline{\theta_t}\|^2\,\mathrm{d}t\\ &-\sum_{\beta\geqslant 1}\langle\overline{\theta_t}, DU_\beta(x_t)\overline{\theta_t}\rangle^2\,\mathrm{d}t + \sum_{\beta\geqslant 1}\langle\overline{\theta_t}, DU_\beta(x_t)\overline{\theta_t}\rangle\,\mathrm{d}W_t^\beta.\end{aligned} \tag{27}$$

We see that the process $\{(x_t,\theta_t) : t \geqslant 0\}$ given by (9) and (26) is also a diffusion process on $\mathbf{R}^2\times\mathbf{S}^1$. If μ is an invariant probability measure for x_t on $\mathbf{R}^2$ then there is at least one invariant probability measure v, say, for (x_t,θ_t) on $\mathbf{R}^2\times\mathbf{S}^1$ with $\mathbf{R}^2$ marginal μ. In the same way that μ is rotation invariant on $\mathbf{R}^2$, the measure v, if unique, must be invariant under the rotations $(x,\theta)\mapsto(R_s x,\theta+s)$ of $\mathbf{R}^2\times\mathbf{S}^1$.

Theorem 4.2. *Assume* A2, A3 *and* A4.

(i) *Suppose the diffusion* (x_t,θ_t) *on* $\mathbf{R}^2\times\mathbf{S}^1$ *given by* (9) *and* (26) *has at most one invariant probability measure. Then it has a unique invariant probability measure* v, *say, and* v_ε *converges weakly to* v *as* $\varepsilon\to 0$. *Moreover, if* $F(x,\theta)$ *is any continuous function on* $\mathbf{R}^2\times\mathbf{S}^1$ *such that* $F(x,\theta)/G(x)\to 0$ *as* $\|x\|\to\infty$ *then* $F\in L^1(v)$ *and*

$$\int F(x,\theta)\,\mathrm{d}v_\varepsilon(x,\theta)\to\int F(x,\theta)\,\mathrm{d}v(x,\theta)$$

as $\varepsilon\to 0$.

(ii) *Assume additionally that*

$$\lim_{\|x\|\to\infty}\frac{\|DV_0(x)\|+\sum_{\alpha=1}^k\|DV_\alpha(x)\|^2}{G(x)}=0. \tag{28}$$

The almost sure Lyapunov exponent for system (9)

$$\tilde{\lambda}=\lim_{t\to\infty}\frac{1}{t}\log\|v_t\|$$

exists with probability 1 *for* μ*-almost all* x_0 *and all* $v_0\neq 0$, *and is given by the formula*

$$\tilde{\lambda}=\int Q_U(x,\theta)\,\mathrm{d}v(x,\theta), \tag{29}$$

where

$$Q_U(x,\theta)=\langle\bar{\theta},DU_0(x)\bar{\theta}\rangle+\sum_{\beta\geqslant 1}\left(\frac{1}{2}\|DU_\beta(x)\bar{\theta}\|^2-\langle\bar{\theta},DU_\beta(x)\bar{\theta}\rangle^2\right).$$

(iii) *In addition to the assumptions in* (i) *and* (ii) *suppose that the process* $\{(x_t^\varepsilon,\theta_t^\varepsilon) : t\geqslant 0\}$ *on* $\mathbf{R}^2\times\mathbf{S}^1$ *given by* (7) *and* (20) *has at most one invariant probability measure, for sufficiently small* $\varepsilon>0$. *Then*

$$\tilde{\lambda}_\varepsilon\sim\varepsilon^2\tilde{\lambda}\quad\textit{as }\varepsilon\to 0.$$

Proof. To prove (i) we extend the methods used in the proofs of Theorems 3.1 and 4.1. Define φ_t and $\tilde{x}_t^\varepsilon$ as in (16) and (17) and define $\tilde{v}_t^\varepsilon=R_{-\varphi_t}v_t^\varepsilon$. Then the direction (in polar coordinates) of $\tilde{v}_t^\varepsilon$ is $\tilde{\theta}_t^\varepsilon=\theta_t^\varepsilon-\varphi_t$. The method of proof of Theorem 3.1 can easily be adapted to show that the process $\{(\tilde{x}_{t/\varepsilon^2}^\varepsilon,\tilde{\theta}_{t/\varepsilon^2}^\varepsilon) : 0\leqslant t\leqslant T\}$ converges in distribution to the process $\{(x_t,\theta_t) : 0\leqslant t\leqslant T\}$ given by (9) and (26). This involves an application of Theorem A.2 to the process $\{(\varphi_t,\tilde{x}_t^\varepsilon,\tilde{\theta}_t^\varepsilon) : t\geqslant 0\}$; the calculation of the generator for the limit process uses Eqs. (10)–(14). Furthermore, the tightness of the μ_ε given by Lemma 4.1 implies tightness of the family v_ε. The proof of Theorem 4.1 can now be extended, using the measures $\tilde{v}_\varepsilon$ on $\mathbf{S}^1 \times \mathbf{R}^2 \times \mathbf{S}^1$ defined by

$$\int f(\varphi,x,\theta)\,\mathrm{d}\tilde{v}_e(\varphi,x,\theta) = \int f(\varphi,R_\varphi x,\theta+\varphi)\,\mathrm{d}v_\varepsilon(x,\theta)\,\mathrm{d}m(\varphi).$$

Theorem A.3 can be applied to give the existence of the unique v, and also the weak convergence of $\tilde{v}_\varepsilon$ to $m \times v$. Then, as in the proof of Theorem 4.1, we get

$$\int h(x,\theta)\,\mathrm{d}v_\varepsilon(x,\theta) \to \int \left(\int h(R_\varphi x,\theta+\varphi)\,\mathrm{d}v(x,\theta) \right) \mathrm{d}m(\varphi)$$

and the weak convergence $v_\varepsilon \Rightarrow v$ follows using the rotation invariance of v. The proof of (i) is completed using the uniform integrability estimate from Lemma 4.1. To prove (ii) notice first that

$$\|DU_0(x)\| \leqslant M_s\{\|DV_0(R_s x)\|\}$$

and

$$\sum_{\beta\geqslant 1} \|DU_\beta(x)\|^2 \leqslant M_s \left\{ \sum_{\alpha=1}^{k} \|DV_\alpha(R_s x)\|^2 \right\}.$$

Since we may assume that $G(x)$ has rotational symmetry, condition (28) implies that

$$\lim_{\|x\|\to\infty} \frac{\|DU_0(x)\| + \sum_{\beta\geqslant 1} \|DU_\beta(x)\|^2}{G(x)} = 0.$$

The method of proof of Proposition 4.1 can now be applied to complete the proof of (ii). To prove (iii) we first take $F(x,\theta) = Q_V(x,\theta)$ in part (i) to obtain

$$\int Q_V(x,\theta)\,\mathrm{d}v_\varepsilon(x,\theta) \to \int Q_V(x,\theta)\,\mathrm{d}v(x,\theta)$$

as $\varepsilon \to 0$. Since the measure v is unique, it is invariant under rotations $(x,\theta) \mapsto (R_s x, \theta+s)$ of $\mathbf{R}^2 \times \mathbf{S}^1$, and so

$$\int Q_V(x,\theta)\,\mathrm{d}v(x,\theta) = \int Q_V(R_s x,\theta+s)\,\mathrm{d}v(x,\theta)$$

for any s. Now we average over the rotation parameter s and use Fubini's theorem to obtain

$$\int Q_V(x,\theta)\,\mathrm{d}v(x,\theta) = M_s \left\{ \int Q_V(R_s x,\theta+s)\,\mathrm{d}v(x,\theta) \right\}$$

$$= \int M_s\{Q_V(R_s x,\theta+s)\}\,\mathrm{d}v(x,\theta).$$

Recalling the definitions of $Q_V(x,\theta)$ and $Q_U(x,\theta)$ given below (24) and (29), and using relations (11) and (14), we obtain

$$M_s\{Q_V(R_s x,\theta+s)\} = Q_U(x,\theta).$$

Together we have

$$\begin{aligned}\int Q_V(x,\theta)\,\mathrm{d}\nu_\varepsilon(x,\theta) \to \int Q_V(x,\theta)\,\mathrm{d}\nu(x,\theta)\\ = \int M_s\{Q_V(R_s x,\theta+s)\}\,\mathrm{d}\nu(x,\theta)\\ = \int Q_U(x,\theta)\,\mathrm{d}\nu(x,\theta).\end{aligned}$$

Finally the assumptions of uniqueness of the measures ν and ν_ε imply that we can use formulas (24) and (29), so that

$$\frac{1}{\varepsilon^2}\tilde{\lambda}_\varepsilon = \int Q_V(x,\theta)\,\mathrm{d}\nu_\varepsilon(x,\theta) \to Q_U(x,\theta)\,\mathrm{d}\nu(x,\theta) = \tilde{\lambda}$$

as $\varepsilon \to 0$, and we are done. □

Remark. The calculation of $\tilde{\lambda}_\varepsilon$ using (24) is a 3-dimensional problem. The rotational symmetries in ν and $Q_U(x,\theta)$ imply that the evaluation of $\tilde{\lambda}$ using (29) is a 2-dimensional problem.

5. Multiplicative noise

Throughout this section we will assume that $V_\alpha(0)=0$ for all $\alpha \geqslant 0$. Then 0 is a fixed point for the one point motions of both (7) and (9). We will study the behavior of systems (7) and (9) on the state space $\mathbf{R}^2 \setminus \{0\}$. The results in this section will be similar to those in Section 4, except that the invariant probability measures μ_ε and μ will be on $\mathbf{R}^2 \setminus \{0\}$ and the invariant probability measures ν_ε and ν will be on $(\mathbf{R}^2 \setminus \{0\}) \times \mathbf{S}^1$.

We begin by studying the behavior of the one-point motions of (7) and (9) near 0. Linearizing the original system (7) at 0 gives the linear stochastic differential equation

$$\mathrm{d}u_t^\varepsilon = -\omega J u_t^\varepsilon\,\mathrm{d}t + \varepsilon^2 DV_0(0)u_t^\varepsilon\,\mathrm{d}t + \varepsilon\sum_{\alpha=1}^{k} DV_\alpha(0)u_t^\varepsilon\,\mathrm{d}W_t^\alpha. \tag{30}$$

Notice that this is simply equation (19) with $x_t^\varepsilon \equiv 0$. The (top) almost sure Lyapunov exponent λ^ε and moment Lyapunov exponent $g^\varepsilon(p)$ for (30) are defined by

$$\lambda^\varepsilon = \lim_{t\to\infty}\frac{1}{t}\log\|u_t^\varepsilon\| \quad \text{and} \quad g^\varepsilon(p) = \lim_{t\to\infty}\frac{1}{t}\log\mathbf{E}\|u_t^\varepsilon\|^p,$$

where the limits exist and do not depend on the non-zero initial vector u_0^ε. Similarly we can linearize (9) to obtain

$$\mathrm{d}u_t = DU_0(0)u_t + \sum_{\beta\geqslant 1} DU_\beta(0)u_t\,\mathrm{d}W_t^\beta \tag{31}$$

with its corresponding almost sure Lyapunov exponent λ and moment Lyapunov exponent $g(p)$. Applying Itô's formula to (31) and making use of the rotational symmetry in $DU_0(0)$ and the $DU_\beta(0)$, $\beta \geqslant 1$, we get

$$\mathrm{d}\log\|u_t\| = \langle e, DU_0(0)e\rangle\,\mathrm{d}t + \sum_{\beta\geqslant 1}\left(\frac{1}{2}\|DU_\beta(0)e\|^2 - \langle e, DU_\beta(0)e\rangle^2\right)\mathrm{d}t$$

$$+\sum_{\beta\geqslant 1}\langle e, DU_\beta(0)e\rangle\,\mathrm{d}W_t^\beta,$$

where $e = [\begin{smallmatrix}1\\0\end{smallmatrix}]$. Therefore

$$\lambda = \langle e, DU_0(0)e\rangle + \sum_{\beta\geqslant 1}\left(\frac{1}{2}\|DU_\beta(0)e\|^2 - \langle e, DU_\beta(0)e\rangle^2\right) = Q_U(0,0) \tag{32}$$

and

$$g(p) = \lambda p + \frac{1}{2}\bar{\sigma}^2 p^2, \tag{33}$$

where $\bar{\sigma}^2 = \sum_{\beta\geqslant 1}\langle e, DU_\beta(0)e\rangle^2$. Notice that assumption A1 implies that $\bar{\sigma}^2 > 0$.

The following result is a direct application of results of Auslender and Milstein [6] (see also Pardoux and Wihstutz [24]) for the almost sure Lyapunov exponent and Khasminskii and Moshchuk [18] for the moment Lyapunov exponent.

Proposition 5.1. *Assume* A1. *Then as* $\varepsilon \to 0$

$$\lambda^\varepsilon \sim \varepsilon^2\lambda$$

$$g^\varepsilon(p) \sim \varepsilon^2 g(p).$$

We go on to consider the relationship between the Lyapunov exponents λ^ε, λ and the behavior near 0 of the diffusions with generators L_ε, $\bar{L}$. First we need a technical lemma.

Lemma 5.1. *Assume* A1 *and* $\lambda > 0$, *and let* $0 < \gamma < 2\lambda/\bar{\sigma}^2$. *Then there exists* $c < \infty$ *and* $0 < r_0 < r_1 < \infty$ *and* $\varepsilon_0 > 0$ *such that for each* $0 < \varepsilon \leqslant \varepsilon_0$ *there exists a* C^2 *positive function* $f_\varepsilon(x)$ *on* $\mathbf{R}^2 \setminus \{(0,0\}$ *with the properties*

$$\sup_{r_0\leqslant\|x\|\leqslant r_1}\left(|f_\varepsilon(x)| + \sum_{\alpha=1}^{k}|(V_\alpha.f_\varepsilon)(x)|\right) \leqslant c$$

and

$$L_\varepsilon f_\varepsilon(x) \leqslant -\varepsilon^2\|x\|^{-\gamma} \quad \textit{whenever } 0 < \|x\| \leqslant r_1.$$

Remark. The value $-2\lambda/\bar{\sigma}^2$ is the stability index (see Arnold and Khasminskii [2]) for system (9).

Proof. Let $T\bar{L}$ denote the generator for the linearized process $\{u_t : t \geqslant 0\}$ given by (31). Then

$$T\bar{L}(\|u\|^{-\gamma}) = g(-\gamma)\|u\|^{-\gamma}.$$

Our choice of γ ensures that $g(-\gamma)<0$. Let TL denote the generator of the process $\{v_t : t \geqslant 0\}$ given by

$$\mathrm{d}v_t = DV_0(0)v_t\,\mathrm{d}t + \sum_{\alpha=1}^{k} DV_\alpha(0)v_t\,\mathrm{d}W_t^\alpha.$$

Then

$$TL(\|v\|^{-\gamma}) = F(\theta)\|v\|^{-\gamma}$$

for some smooth function $F:\mathbf{S}^1 \to \mathbf{R}$, where we use polar coordinates $v=\|v\|[\begin{smallmatrix}\cos\theta\\ \sin\theta\end{smallmatrix}]$. Analogously to the way that $\bar{L}$ was obtained from L in (8) we have

$$T\bar{L}h(v) = \frac{1}{2\pi}\int_0^{2\pi} TL(h\circ R_{-s})(R_s v)\,\mathrm{d}s.$$

This can be verified directly using the symmetry relations (11) and (14). Taking $h(v)=\|v\|^{-\gamma}$ we get

$$g(-\gamma) = \frac{1}{2\pi}\int_0^{2\pi} F(s)\,\mathrm{d}s$$

and so there exists smooth $\tilde{F}:\mathbf{S}^1\to\mathbf{R}$ such that $\omega\tilde{F}'(\theta)=g(-\gamma)-F(\theta)$. Now we switch from the linearization TL to L, and switch from the variable v to the variable $x\in\mathbf{R}^2\setminus\{0\}$ with polar coordinates r and ϕ. Define $f_\varepsilon(x)=r^{-\gamma}(1+\varepsilon^2\tilde{F}(\phi))$. The condition $f_\varepsilon(x)\geqslant 0$ will be satisfied so long as ε_0 is chosen to satisfy $\varepsilon_0 \leqslant 1/\sup_\phi|\tilde{F}(\phi)|$. We obtain

$$\begin{aligned} L_\varepsilon f_\varepsilon(x) &= \varepsilon^2 r^{-\gamma}\omega\tilde{F}'(\phi) + \varepsilon^2 L(r^{-\gamma}) + \varepsilon^4 L(r^{-\gamma}\tilde{F}(\phi)) \\ &= \varepsilon^2 r^{-\gamma} g(-\gamma) + \varepsilon^2[L(r^{-\gamma}) - TL(r^{-\gamma})] + \varepsilon^4 L(r^{-\gamma}\tilde{F}(\phi)). \end{aligned}$$

Now by Baxendale and Stroock [10] Corollary 3.10, see also Arnold and Khasminskii [2], there exists $c_1<\infty$ such that

$$|L(r^{-\gamma}) - TL(r^{-\gamma})| \leqslant c_1 r^{1-\gamma}$$

and

$$|L(r^{-\gamma})\tilde{F}(\phi)| \leqslant c_1 r^{-\gamma}$$

for $0<r<1$. Thus

$$L_\varepsilon f_\varepsilon(x) \leqslant \varepsilon^2 r^{-\gamma}[g(-\gamma) + c_1(r+\varepsilon^2)]$$

for $0<r<1$. Since $g(-\gamma)<0$ there exist $r_1\in(0,1)$ and $\varepsilon_0>0$ such that such $g(-\gamma)+c_1(r+\varepsilon^2)\leqslant(1/2)g(-\gamma)<0$ for $r\leqslant r_1$ and $\varepsilon\leqslant\varepsilon_0$. The result now follows by scaling f_ε by a factor $-2/g(-\gamma)$. Notice that r_0 can be chosen arbitrarily in $(0,r_1)$ and then the existence of c is guaranteed by the explicit form of f_ε. □

Proposition 5.2. *Assume* A1, A2 *and* A4.

(i) *If* $\lambda<0$ *then for all sufficiently small* $\varepsilon>0$ *the diffusion* $\{x_t^\varepsilon : t\geqslant 0\}$ *with generator* L_ε *satisfies*

$$\mathbf{P}^x(\|x_t^\varepsilon\|\to 0 \text{ as } t\to\infty) \to 1 \text{ as } x\to 0.$$

(ii) *If* $\lambda > 0$ *there is* $\varepsilon_0 > 0$ *such that for* $0 < \varepsilon \leqslant \varepsilon_0$ *there exists an invariant probability measure* μ_ε *on* $\mathbf{R}^2 \setminus \{0\}$ *for the diffusion* L_ε. *Moreover, if* $0 < \gamma < 2\lambda/\bar{\sigma}^2$ *there exists* $\varepsilon_1 \in (0, \varepsilon_0]$ *and* $c_1 < \infty$ *such that any invariant probability measure* μ_ε *for* $0 < \varepsilon \leqslant \varepsilon_1$ *satisfies*

$$\int (\|x\|^{-\gamma} + G(x))\, \mathrm{d}\mu_\varepsilon(x) \leqslant c_1.$$

Proof. The result (i) is a direct consequence of Proposition 5.1 and the local stable manifold theorem (see Ruelle [25]). To prove (ii) we adapt the proof of Lemma 4.1. Let f_ε and $0 < r_0 < r_1$ be as in Lemma 5.1, and let F_ε and $0 < R_1 < R_2$ be as in assumption A4. It is clear from the proof of Lemma 5.1 that we may assume $r_1 < R_1$. We can assume the same ε_0 and c. Let $h_1(x)$ and $h_2(x)$ be smooth functions $\mathbf{R}^2 \to [0, 1]$ so that $h_1(x) = 1$ for $\|x\| \leqslant r_0$, and $h_1(x) = 0$ for $\|x\| \geqslant r_1$, and $h_2(x) = 0$ for $\|x\| \leqslant R_1$, and $h_2(x) = 1$ for $\|x\| \geqslant R_2$. Define the function

$$\tilde{F}_\varepsilon(x) = h_1(x) f_\varepsilon(x) + h_2(x) F_\varepsilon(x).$$

Then $\tilde{F}_\varepsilon(x) \geqslant 0$ for $\|x\| \neq 0$ and (arguing as in the proof of Lemma 4.1)

$$L_\varepsilon \tilde{F}_\varepsilon(x) \leqslant -\varepsilon^2 [h_1(x)\|x\|^{-\gamma} + h_2(x) G(x)] + \varepsilon^2 c_1$$

for some $c_1 < \infty$. The remainder of the proof is repeat of the methods used in the proof of Lemma 4.1. □

Remark. Suppose in Proposition 5.2 we make the extra assumption that for each $r > 0$ there exists $\alpha \geqslant 1$ and x with $\|x\| = r$ so that $\langle x, V_\alpha(x) \rangle \neq 0$. Then in (i) we can apply Theorem 2.12 of Baxendale [7] and assert that for small enough $\varepsilon > 0$

$$\mathbf{P}^x(\|x_t^\varepsilon\| \to 0 \quad \text{as } t \to \infty) = 1$$

for all $x \neq 0$. In (ii) we can apply Theorem 5 of Arnold and Kliemann [3] and assert that μ_ε is unique.

We have obtained results on the existence and tightness of invariant probability measures μ_ε for the one point motion of (7) on $\mathbf{R}^2 \setminus \{0\}$. The rest of this section is now essentially a repeat of Section 4 with $\mathbf{R}^2$ replaced $\mathbf{R}^2 \setminus \{0\}$ and using Proposition 5.2 in place of Lemma 4.1.

Theorem 5.1. *Assume* A1, A2, A3 *and* A4, *and* $\lambda > 0$. *Suppose that the diffusion with generator* $\bar{L}$ *has at most one invariant probability measure on* $\mathbf{R}^2 \setminus \{0\}$. *Then it has a unique invariant probability measure* μ, *say, on* $\mathbf{R}^2 \setminus \{0\}$ *and* μ_ε *converges weakly to* μ *as* $\varepsilon \to 0$. *Moreover, if* $f(x)$ *is any continuous function on* $\mathbf{R}^2 \setminus \{0\}$ *such that* $f(x)/G(x) \to 0$ *as* $\|x\| \to \infty$ *and* $f(x)\|x\|^\gamma \to 0$ *as* $\|x\| \to 0$ *for some* $\gamma < 2\lambda/\bar{\sigma}^2$ *then* $f \in L^1(\mu)$ *and*

$$\int f(x)\, \mathrm{d}\mu_\varepsilon(x) \to \int f(x)\, \mathrm{d}\mu(x)$$

as $\varepsilon \to 0$.

Proof. The proof is essentially the same as for Theorem 4.1. The only essential difference is that we work with $\mathbf{R}^2 \setminus \{0\}$ in place of $\mathbf{R}^2$. Proposition 5.2 gives the required tightness condition for the μ_ε. In the proof of Theorem A.3, if the $\tilde{\mu}_e$ are tight as probability measures on $\mathbf{S}^1 \times (\mathbf{R}^2 \setminus \{0\})$, then any weak limit $\tilde{\mu}$ of a subsequence of them is also a probability measure on $\mathbf{S}^1 \times (\mathbf{R}^2 \setminus \{0\})$. The method of proof of Theorem A.3 implies that $\tilde{\mu} = m \times \mu$ for some invariant probability measure μ on $\mathbf{R}^2 \setminus \{0\}$, and the uniqueness of μ on $\mathbf{R}^2 \setminus \{0\}$ completes the argument. □

We now consider the linearizations of systems (7) and (9) along trajectories. These are given by Eqs. (19)–(21) and (25)–(27) in Section 4. By Proposition 5.2, for $\lambda > 0$ and sufficiently small $\varepsilon > 0$ there exists an invariant probability measure μ_ε for x_t^ε on $\mathbf{R}^2 \setminus \{0\}$, and hence there exists also an invariant probability measure v_ε for $(x_t^\varepsilon, \theta_t^\varepsilon)$ on $(\mathbf{R}^2 \setminus \{0\}) \times \mathbf{S}^1$.

Proposition 5.3. *Assume* A1, A2 *and* A4, *and* $\lambda > 0$. *Assume also that*

$$\limsup_{\|x\| \to \infty} \frac{\|DV_0(x)\| + \sum_{\alpha=1}^k \|DV_\alpha(x)\|^2}{G(x)} < \infty,$$

and that the process $\{(x_t^\varepsilon, \theta_t^\varepsilon) : t \geqslant 0\}$ *on* $(\mathbf{R}^2 \setminus \{0\}) \times \mathbf{S}^1$ *given by* (7) *and* (20) *has at most one invariant probability measure. Then it has a unique invariant probability measure* v_ε, *say, on* $(\mathbf{R}^2 \setminus \{0\}) \times \mathbf{S}^1$. *Let* μ_ε *denote the* $\mathbf{R}^2 \setminus \{0\}$ *marginal of* v_ε. *The almost sure Lyapunov exponent for system* (7)

$$\tilde{\lambda}_\varepsilon = \lim_{t \to \infty} \frac{1}{t} \log \|v_t^\varepsilon\|$$

exists with probability 1 *for* μ_ε*-almost all* x_0^ε *and all* $v_0^\varepsilon \neq 0$, *and is given by the formula*

$$\tilde{\lambda}_\varepsilon = \varepsilon^2 \int Q_V(x, \theta) \, \mathrm{d}v_\varepsilon(x, \theta),$$

where

$$Q_V(x, \theta) = \langle \bar{\theta}, DV_0(x)\bar{\theta} \rangle + \sum_{\alpha=1}^k \left(\frac{1}{2} \|DV_\alpha(x) R_\theta e\|^2 - \langle \bar{\theta}, DV_\alpha(x)\bar{\theta} \rangle^2 \right).$$

Proof. This result is proved using the same method as in the proof of Proposition 4.1. □

Theorem 5.2. *Assume* A1, A2, A3 *and* A4, *and* $\lambda > 0$.

(i) *Suppose that the diffusion* (x_t, θ_t) *on* $(\mathbf{R}^2/\{0\}) \times \mathbf{S}^1$ *given by* (9) *and* (26) *has at most one invariant probability measure. Then it has a unique invariant probability measure* v, *say, on* $(\mathbf{R}^2/\{0\}) \times \mathbf{S}^1$, *and* v_ε *converges weakly to* v *as* $\varepsilon \to 0$. *Moreover, if* $f(x, \theta)$ *is any continuous function on* $(\mathbf{R}^2 \setminus \{0\}) \times \mathbf{S}^1$ *such that* $f(x, \theta)/G(x) \to 0$ *as* $\|x\| \to \infty$ *and* $f(x, \theta)\|x\|^\gamma \to 0$ *as* $\|x\| \to 0$ *for some*

$\gamma < 2\lambda/\bar{\sigma}^2$ then $f \in L^1(\nu)$ and

$$\int f(x,\theta)\,\mathrm{d}\nu_\varepsilon(x,\theta) \to \int f(x,\theta)\,\mathrm{d}\nu(x,\theta)$$

as $\varepsilon \to 0$.

(ii) *Assume additionally that*

$$\lim_{\|x\|\to\infty} \frac{\|DV_0(x)\| + \sum_{\alpha=1}^{k} \|DV_\alpha(x)\|^2}{G(x)} = 0. \tag{34}$$

The almost sure Lyapunov exponent for system (9)

$$\tilde{\lambda} = \lim_{t\to\infty} \frac{1}{t} \log \|v_t\|$$

exists with probability 1 for μ-almost all x_0 and all $v_0 \neq 0$, and is given by the formula

$$\tilde{\lambda} = \int Q_U(x,\theta)\,\mathrm{d}\nu(x,\theta), \tag{35}$$

where

$$Q_U(x,\theta) = \langle \bar{\theta}, DU_0(x)\bar{\theta}\rangle + \sum_{\beta\geqslant 1} \left(\frac{1}{2} \|DU_\beta(x)R_\theta e\|^2 - \langle \bar{\theta}, Du_\beta(x)\bar{\theta}\rangle^2 \right).$$

(iii) *In addition to the assumptions in* (i) *and* (ii) *suppose that the process $\{(x_t^\varepsilon, \theta_t^\varepsilon) : t \geqslant 0\}$ on $(\mathbf{R}^2 \setminus \{0\}) \times \mathbf{S}^1$ given by* (7) *and* (20) *has at most one invariant probability measure, for sufficiently small $\varepsilon > 0$. Then*

$$\tilde{\lambda}_\varepsilon \sim \varepsilon^2 \tilde{\lambda} \quad \text{as } \varepsilon \to 0.$$

Proof. The proof combines the method of proof used in Theorem 4.2 above with the tightness estimate from Proposition 5.2. The argument involving the restriction to $(\mathbf{R}^2 \setminus \{0\}) \times \mathbf{S}^1$ is essentially the same as in the proof of Theorem 5.1. □

6. Stochastic Duffing–van der Pol equation

In this section we return to the system

$$\begin{cases} \mathrm{d}x_t = -\omega y_t\,\mathrm{d}t \\ \mathrm{d}y_t = \omega x_t\,\mathrm{d}t + \varepsilon^2 \left(\beta y_t + \dfrac{A}{\omega} x_t^3 - B x_t^2 y_t \right) \mathrm{d}t - \dfrac{\varepsilon\sigma}{\omega} x_t\,\mathrm{d}W_t. \end{cases}$$

Now x and y are scalars and we write $z = [\,{}^{x}_{y}\,]$ and $\tilde{z} = [\,{}^{\tilde{x}}_{\tilde{y}}\,]$. We will carry out the averaging process described in Section 3. Consider first the evaluation of $a(z,\tilde{z})$.

Since $V_1(z) = -\frac{\sigma}{\omega}\left[\begin{smallmatrix} 0 & 0 \\ 1 & 0 \end{smallmatrix}\right] z$ we obtain

$$\begin{aligned} R_{-s}V_1(R_s z) &= -\frac{\sigma}{\omega} R_{-s} \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} R_s z \\ &= -\frac{\sigma}{\omega} \begin{bmatrix} (\sin s)(\cos s) & -\sin^2 s \\ \cos^2 s & -(\sin s)(\cos s) \end{bmatrix} z \\ &= \frac{-\sigma}{2\omega}(J - D_s)z, \end{aligned}$$

where

$$D_s = \begin{bmatrix} \sin 2s & \cos 2s \\ \cos 2s & -\sin 2s \end{bmatrix}.$$

Identifying $u \otimes v \in \mathbf{R}^2 \otimes \mathbf{R}^2$ with the 2×2 matrix uv^{T} we obtain

$$a(z, \tilde{z}) = \frac{\sigma^2}{4\omega^2} M_s\{(J - D_s)z\tilde{z}^{\mathrm{T}}(J^{\mathrm{T}} - D_s^{\mathrm{T}})\}.$$

Now $M_s\{D_s\} = 0$ and $M_s\{D_s^{\mathrm{T}}\} = 0$ and

$$\begin{aligned} M_s\{D_s z\tilde{z}^{\mathrm{T}} D_s^{\mathrm{T}}\} &= M_s\left\{ \begin{bmatrix} \sin 2s & \cos 2s \\ \cos 2s & -\sin 2s \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \begin{bmatrix} \tilde{x} & \tilde{y} \end{bmatrix} \begin{bmatrix} \sin 2s & \cos 2s \\ \cos 2s & -\sin 2s \end{bmatrix} \right\} \\ &= \frac{1}{2} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \begin{bmatrix} \tilde{x} & \tilde{y} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \\ &\quad + \frac{1}{2} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \begin{bmatrix} \tilde{x} & \tilde{y} \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \\ &= \frac{1}{2} K_1 z\tilde{z}^{\mathrm{T}} K_1^{\mathrm{T}} + \frac{1}{2} K_2 z\tilde{z}^{\mathrm{T}} K_2^{\mathrm{T}}. \end{aligned}$$

Therefore

$$a(z, \tilde{z}) = \frac{\sigma^2}{4\omega^2}(Jz)(J\tilde{z})^{\mathrm{T}} + \frac{\sigma^2}{8\omega^2}(K_1 z)(K_1\tilde{z})^{\mathrm{T}} + \frac{\sigma^2}{8\omega^2}(K_2 z)(K_2\tilde{z})^{\mathrm{T}}.$$

In this form we can easily read off the three vector fields U_1, U_2 and U_3. In matrix form we get

$$a(z, \tilde{z}) = \frac{\sigma^2}{8\omega^2} \begin{bmatrix} x\tilde{x} + 3y\tilde{y} & -x\tilde{y} - y\tilde{x} \\ -x\tilde{y} - y\tilde{x} & 3x\tilde{x} + y\tilde{y} \end{bmatrix}.$$

Now we evaluate $U_0(z)$. Using polar coordinates $z = r\left[\begin{smallmatrix} \cos\phi \\ \sin\phi \end{smallmatrix}\right] = rR_\phi e$ we have

$$V_0(z) = \begin{bmatrix} 0 \\ \beta r \sin\phi + (A/\omega) r^3 \cos^3\phi - B r^3 \cos^2\phi \sin\phi \end{bmatrix}$$

and so

$$
\begin{aligned}
U_0(z) &= M_s\{R_{-s}V_0(R_s r R_\phi e)\}\\
&= R_\phi M_s\{R_{-s}V_0(rR_s e)\}\\
&= R_\phi M_s\left\{(\beta r\sin s + (A/\omega)r^3\cos^3 s - Br^3\cos^2 s\sin s)\begin{bmatrix}\sin s\\ \cos s\end{bmatrix}\right\}\\
&= R_\phi\begin{bmatrix}\beta r/2 - Br^3/8\\ 3Ar^3/8\omega\end{bmatrix}\\
&= \left(\frac{\beta}{2}-\frac{Br^2}{8}\right)z - \frac{3Ar^2}{8\omega}Jz.
\end{aligned}
$$

The second line above was obtained by replacing $s+\phi$ with s. Therefore for the stochastic Duffing–van der Pol equation the stochastically averaged system is

$$
\begin{aligned}
\mathrm{d}z_t &= \left(\frac{\beta}{2}-\frac{B\|z_t\|^2}{8}\right)z_t\,\mathrm{d}t - \frac{3A\|z_t\|^2}{8\omega}Jz_t\,\mathrm{d}t\\
&\quad+\frac{\sigma}{2\sqrt{2}\omega}K_1 z_t\,\mathrm{d}W_t^1 + \frac{\sigma}{2\sqrt{2}\omega}K_2 z_t\,\mathrm{d}W_t^2 + \frac{\sigma}{2\omega}Jz_t\,\mathrm{d}W_t^3
\end{aligned}
$$

which is exactly our Eq. (6).

Remark. The paper [5] by Arnold et al. contains a stochastic averaging calculation for a version of the stochastic Duffing–van der Pol equation with a more general mixture of white and colored noise. The one-point motion of (6) is consistent with the one-point motion of the system obtained by [5]. However in [5] the stochastic averaging is done for the one-point motion only. There are many possible stochastic differential equations for which the one-point motion agrees with that of (6). In [5] the arbitrary choice is made to use the stochastic differential equation

$$
\begin{aligned}
\mathrm{d}z_t &= \left(\frac{\beta}{2}-\frac{B\|z_t\|^2}{8}\right)z_t\,\mathrm{d}t - \frac{3A\|z_t\|^2}{8\omega}Jz_t\,\mathrm{d}t\\
&\quad+\frac{\sigma}{2\sqrt{2}\omega}z_t\,\mathrm{d}W_t^1 + \frac{\sqrt{3}\sigma}{2\sqrt{2}\omega}Jz_t\,\mathrm{d}W_t^2. \qquad (36)
\end{aligned}
$$

This system is easier to analyze than (6) since there is more decoupling between the radial and angular parts of the motion. In particular, for system (36) when $\lambda>0$ there is the exact result $\tilde{\lambda}=0$. In contrast, for system (6) we will have to rely on numerical methods to evaluate $\tilde{\lambda}$, see Section 7. However the results of Section 5 tell us that it is the Lyapunov exponent $\tilde{\lambda}$ for (6), not for (36), which gives the correct asymptotic behavior for the Lyapunov exponent $\tilde{\lambda}_\varepsilon$ of the original stochastic Duffing–van der Pol equation.

Clearly system (5) satisfies A1. It is easy to check that it satisfies A2 and A3, using the functions $F_\varepsilon(x,y)=\omega^2(x^2+y^2)/2+\varepsilon^2Ax^4/4$ and $F_0(x,y)=\omega^2(x^2+y^2)/2$. Now we verify the condition A4. Fix $\delta>0$ and define

$$H_\varepsilon(x,y)=\frac{\omega^2-\varepsilon^4\delta^2}{2}x^2+\frac{\varepsilon^2 4A+\varepsilon^4\delta B}{12}x^4+\frac{1}{2}\left(\omega y+\varepsilon^2(\beta+\delta)x-\varepsilon^2\frac{B}{3}x^3\right)^2.$$

Then

$$L_\varepsilon H_\varepsilon(x,y)=\varepsilon^2\left[-\omega^2\delta y^2-\varepsilon^2\frac{AB}{3}x^6-\left(\frac{\omega^2 B}{3}-\varepsilon^2A(\beta+\delta)\right)x^4\right.$$
$$\left.+\left(\frac{\sigma^2}{2}+\omega^2\beta+\omega^2\delta\right)x^2-\varepsilon^2\omega\beta\delta xy\right].$$

Assuming that $B>0$ it is easy to see there exist $\varepsilon_0>0$ and constants K_1, K_2 and R_0 such that

$$H_\varepsilon(x,y)\leqslant K_1(x^6+y^2),$$
$$L_\varepsilon H_\varepsilon(x,y)\leqslant -\varepsilon^2K_2(x^4+y^2),$$

whenever $0<\varepsilon\leqslant\varepsilon_0$ and $\|(x,y)\|\geqslant R_0$. Then on the set $\|(x,y)\|\geqslant R_0$ we have

$$\frac{L_\varepsilon(\exp(\alpha H_\varepsilon^{1/3}))(x,y)}{\exp(\alpha H_\varepsilon(x,y)^{1/3})}$$
$$=\frac{\alpha}{3}H_\varepsilon(x,y)^{-2/3}L_\varepsilon H_\varepsilon(x,y)$$
$$+\left(\frac{\alpha^2}{9}H_\varepsilon(x,y)^{-4/3}-\frac{2\alpha}{9}H_\varepsilon(x,y)^{-5/3}\right)\frac{\varepsilon^2\sigma^2x^2}{\omega^2}\left(\frac{\partial H_\varepsilon(x,y)}{\partial y}\right)^2$$
$$\leqslant\frac{\alpha}{3}H_\varepsilon(x,y)^{-2/3}L_\varepsilon H_\varepsilon(x,y)+\frac{2\alpha^2}{9}\varepsilon^2\sigma^2x^2H_\varepsilon(x,y)^{-1/3}$$
$$\leqslant-\frac{\varepsilon^2\alpha}{3H_\varepsilon(x,y)^{2/3}}\left[K_2(x^4+y^2)-\frac{2K_1^{1/3}\alpha\sigma^2}{3}x^2(x^6+y^2)^{1/3}\right].$$

Therefore if α is sufficiently small (depending only on K_1, K_2, σ^2 and R_0) we obtain the estimate

$$L_\varepsilon(\exp(\alpha H_\varepsilon^{1/3}))(x,y)\leqslant-\varepsilon^2K_3\exp(\alpha H_\varepsilon(x,y)^{1/3})$$

for $x^2+y^2\geqslant R_1$ for some constants K_3 and R_1. Finally, since

$$\frac{\omega^2}{2}(x^2+y^2)\leqslant\frac{\omega^2}{2}x^2+\left(\omega y+\varepsilon^2(\beta+\delta)x-\frac{\varepsilon^2B}{3}x^3\right)^2+\varepsilon^4\left(\frac{B}{3}x^3-(\beta+\delta)x\right)^2$$
$$\leqslant\left(\frac{\omega^2}{2}+2\varepsilon^4(\beta+\delta)^2\right)x^2+\left(\omega y+\varepsilon^2(\beta+\delta)x-\frac{\varepsilon^2B}{3}x^3\right)^2$$
$$+\varepsilon^4\frac{2B^2}{9}x^6$$

it follows that there are positive constants c_1 and c_2 such that $x^2 + y^2 \leqslant c_1 H_\varepsilon(x, y) + c_2 H_\varepsilon(x, y)^{3/2}$ for all $\varepsilon \leqslant \varepsilon_0$, for sufficiently small ε_0. In particular $x^2 + y^2 \leqslant (c_1 + c_2) H_\varepsilon(x, y)^{3/2}$ whenever $x^2 + y^2 \geqslant c_1 + c_2$. Replacing R_1 with $\max(R_1, c_1 + c_2)$ and α with $\alpha/(c_1 + c_2)$ we have the estimate

$$L_\varepsilon(\exp(\alpha H_\varepsilon^{1/3}))(x, y) \leqslant -\varepsilon^2 K_3 \exp(\alpha(x^2 + y^2)^{2/9})$$

whenever $\varepsilon \leqslant \varepsilon_0$ and $x^2+y^2 \geqslant R_1^2$. Therefore A4 is satisfied with $G(x, y)=K_3 \exp(\alpha(x^2+y^2)^{2/9})$ and $F_\varepsilon(x) = \exp(\alpha H_\varepsilon(x, y)^{1/3})$.

We have verified conditions A1–A4 for Eqs. (5) and (6). Moreover, since $G(x, y)$ grows faster than any polynomial, it is clear that condition (34) in Theorem 5.2 is satisfied.

From Eqs. (32) and (33) we obtain

$$\lambda = \frac{\beta}{2} + \frac{\sigma^2}{8\omega^2} \quad \text{and} \quad g(p) = \left(\frac{\beta}{2} + \frac{\sigma^2}{8\omega^2}\right) p + \frac{\sigma^2 p^2}{16\omega^2}.$$

Part (i) of Theorem 1.1 follows by Proposition 5.1. Writing (6) in polar coordinates $z = r\left[\begin{smallmatrix}\cos\phi \\ \sin\phi\end{smallmatrix}\right]$ we obtain

$$\mathrm{d}r_t = \left(\frac{\beta r_t}{2} - \frac{B r_t^3}{8} + \frac{3\sigma^2 r_t}{16\omega^2}\right) \mathrm{d}t + \frac{\sigma}{2\sqrt{2}\omega} r_t \cos 2\phi_t \, \mathrm{d}W_t^1 + \frac{\sigma}{2\sqrt{2}\omega} r_t \sin 2\phi_t \, \mathrm{d}W_t^2 \tag{37}$$

and

$$\mathrm{d}\phi_t = \frac{A r_t^2}{8\omega} \mathrm{d}t - \frac{\sigma}{2\sqrt{2}\omega} \sin 2\phi_t \, \mathrm{d}W_t^1 + \frac{\sigma}{2\sqrt{2}\omega} \cos 2\phi_t \, \mathrm{d}W_t^2 - \frac{\sigma}{2\omega} \mathrm{d}W_t^3. \tag{38}$$

The uniqueness of any invariant measure μ on $\mathbf{R}^2 \setminus \{0\}$ follows immediately. Moreover from (37) we see that the process r_t is a diffusion with generator

$$\mathscr{L} = \frac{\sigma^2 r^2}{16\omega^2} \frac{\mathrm{d}^2}{\mathrm{d}r^2} + \left(\frac{\beta r}{2} + \frac{3\sigma^2 r}{16\omega^2} - \frac{B r^3}{8}\right) \frac{\mathrm{d}}{\mathrm{d}r}.$$

So, when $\lambda = \beta/2 + \sigma^2/8\omega^2 > 0$ the process r_t has invariant measure ρ, say, with density

$$c r^{1+8\omega^2\beta/\sigma^2} \exp\left(-\frac{B\omega^2 r^2}{\sigma^2}\right)$$

with respect to Lebesgue measure on $(0, \infty)$, where the normalizing constant c is given $c^{-1} = 2(\frac{\sigma^2}{B\omega^2})^{\omega^2\beta/2\sigma^2 + 1/8} \Gamma(\frac{\omega^2\beta}{2\sigma^2} + \frac{1}{8})$. Then the invariant measure μ on $\mathbf{R}^2 \setminus \{0\}$ has density

$$\frac{c}{2\pi} r^{8\omega^2\beta/\sigma^2} \exp\left(-\frac{B\omega^2 r^2}{\sigma^2}\right)$$

with respect to 2-dimensional Lebesgue measure, where $r = \sqrt{x^2 + y^2}$. Part (ii) of Theorem 1.1 follows by Theorem 5.1.

Eqs. (26) and (27) for the linearization of (5) become

$$\mathrm{d}\theta_t = \left(\frac{Ar_t^2}{8\omega} + \frac{Br_t^2}{8}\sin 2(\theta_t - \phi_t) + \frac{3Ar_t^2}{4\omega}\cos^2(\theta_t - \phi_t)\right)\mathrm{d}t$$
$$- \frac{\sigma}{2\sqrt{2}\omega}\sin 2\theta_t\,\mathrm{d}W_t^1 + \frac{\sigma}{2\sqrt{2}\omega}\cos 2\theta_t\,\mathrm{d}W_t^2 - \frac{\sigma}{2\omega}\,\mathrm{d}W_t^3 \tag{39}$$

and

$$\mathrm{d}\log\|v_t\| = \left(\frac{\beta}{2} + \frac{\sigma^2}{8\omega^2} - \frac{Br_t^2}{8} - \frac{Br_t^2}{4}\cos^2(\theta_t - \phi_t) + \frac{3Ar_t^2}{8\omega}\sin 2(\theta_t - \phi_t)\right)\mathrm{d}t$$
$$+ \frac{\sigma}{2\sqrt{2}\omega}\cos 2\theta_t\,\mathrm{d}W_t^1 + \frac{\sigma}{2\sqrt{2}\omega}\sin 2\theta_t\,\mathrm{d}W_t^2. \tag{40}$$

From Eqs. (37)–(39) we see that the generator for the diffusion $\{(r_t, \phi_t, \theta_t) : t \geqslant 0\}$ on $(\mathbf{R}^2 \setminus \{0\}) \times \mathbf{S}^1$ is elliptic at all points where $\phi \neq \theta$. Since $\mathrm{d}(\theta_t - \phi_t) = 3Ar_t^2/4\omega\,\mathrm{d}t$ whenever $\theta_t = \phi_t$, it follows that the diffusion $\{(r_t, \phi_t, \theta_t)\colon t \geqslant 0\}$ has at most one invariant measure ν on $(\mathbf{R}^2 \setminus \{0\}) \times \mathbf{S}^1$. (Here is where we use the assumption that $A > 0$.) Part (iii) of Theorem 1.1 follows by Theorem 5.2.

We proceed with the calculation of the Lyapunov exponent $\tilde{\lambda}$ for the averaged system (6). Let $\psi_t = \theta_t - \phi_t$, so that ψ_t gives the angle between the vectors $\left[\begin{smallmatrix} x_t \\ y_t \end{smallmatrix}\right]$ and v_t. Define new Brownian motion processes

$$\mathrm{d}B_t^1 = (\cos 2\phi_t)\,\mathrm{d}W_t^1 + (\sin 2\phi_t)\,\mathrm{d}W_t^2,$$

$$\mathrm{d}B_t^2 = -(\sin 2\phi_t)\,\mathrm{d}W_t^1 + (\cos 2\phi_t)\,\mathrm{d}W_t^2.$$

Then B_t^1 and B_t^2 are independent standard Brownian motions and the equations for r_t, ψ_t and $\log\|v_t\|$ become

$$\mathrm{d}r_t = \left(\frac{\beta r_t}{2} - \frac{Br_t^3}{8} + \frac{3\sigma^2 r_t}{16\omega^2}\right)\mathrm{d}t + \frac{\sigma}{2\sqrt{2}\omega} r_t\,\mathrm{d}B_t^1 \tag{41}$$

and

$$\mathrm{d}\psi_t = \left(\frac{B}{8} r_t^2 \sin 2\psi_t + \frac{3Ar_t^2}{4\omega}\cos^2\psi_t\right)\mathrm{d}t$$
$$- \frac{\sigma}{2\sqrt{2}\omega}\sin 2\psi_t\,\mathrm{d}B_t^1 + \frac{\sigma}{2\sqrt{2}\omega}(-1 + \cos 2\psi_t)\,\mathrm{d}B_t^2 \tag{42}$$

and

$$\mathrm{d}\log\|v_t\| = \left(\frac{\beta}{2} + \frac{\sigma^2}{8\omega^2} - \frac{Br_t^2}{8} - \frac{Br_t^2}{4}\cos^2\psi_t + \frac{3Ar_t^2}{8\omega}\sin 2\psi_t\right)\mathrm{d}t$$
$$+ \frac{\sigma}{2\sqrt{2}\omega}\cos 2\psi_t\,\mathrm{d}B_t^1 + \frac{\sigma}{2\sqrt{2}\omega}\sin 2\psi_t\,\mathrm{d}B_t^2. \tag{43}$$

These three equations display explicitly the facts that r_t is a diffusion, that the pair $(r_t, \theta_t - \phi_t)$ is a diffusion, and that the drift term Q_U in the equation for $\log\|v_t\|$ is a function of r and $\theta - \phi$ only. All of these facts are direct consequences of the rotational symmetry in the averaged system (6).

Suppose $\lambda = \beta/2 + \sigma^2/8\omega^2 > 0$, and let $\bar{v}$ be the unique invariant probability measure on $(0, \infty)$ for the process $\{(r_t, \psi_t) : t \geqslant 0\}$ given by (41), (42). The existence of $\bar{v}$ follows from the existence of its $(0, \infty)$ marginal ρ, and its uniqueness is proved using the same arguments as for the uniqueness of v above. From Eq. (43) we obtain

$$\tilde{\lambda} = \int Q(r, \psi) \, \mathrm{d}\bar{v}(r, \psi), \tag{44}$$

where

$$Q(r, \psi) = \frac{\beta}{2} + \frac{\sigma^2}{8\omega^2} - \frac{Br^2}{8} - \frac{Br^2}{4} \cos^2 \psi + \frac{3Ar^2}{8\omega} \sin 2\psi \ .$$

Notice that this formula can equally well be obtained from (35) using the rotational symmetry of the integrand Q_U and the measure v. From our explicit formula for ρ we can calculate

$$\int \frac{Br^2}{8} \, \mathrm{d}\rho(r) = \frac{\beta}{2} + \frac{\sigma^2}{8\omega^2} = \lambda.$$

Therefore we obtain

$$\tilde{\lambda} = -\lambda + \int \bar{Q}(r, \psi) \, \mathrm{d}\bar{v}(r, \psi),$$

where

$$\bar{Q}(r, \psi) = -\frac{Br^2}{8} \cos 2\psi + \frac{3Ar^2}{8\omega} \sin 2\psi \ .$$

Our stochastic averaging method has obtained the integral formula above for $\tilde{\lambda}$ which involves just two variables r and ψ, rather than the three original variables r, ϕ and θ. We do not have an exact expression for $\tilde{\lambda}$. In the next section we will present the results of numerical calculations for $\tilde{\lambda}$.

In the original stochastic Duffing–van der Pol equation (4), in addition to ε there are five constants: β, σ, ω, A and B. In the averaged system (6) we have just four constants: β, σ/ω, A/ω and B. Let us consider how the Lyapunov exponents λ and $\tilde{\lambda}$ for (6) depend upon these constants. The first result is simple. We have the explicit formula

$$\lambda = \lambda_{\text{ave}} \left(\beta, \frac{\sigma}{\omega} \right) = \frac{\beta}{2} + \frac{\sigma^2}{8\omega^2}.$$

Now suppose $\lambda > 0$ and write

$$\tilde{\lambda} = \tilde{\lambda}_{\text{ave}} \left(\beta, \frac{\sigma}{\omega}, \frac{A}{\omega}, B \right).$$

Rescaling (6) in time and space (by putting $\tilde{z}_t = C z_{Dt}$ for suitable C and D) we obtain

$$\tilde{\lambda}_{\text{ave}} \left(\beta, \frac{\sigma}{\omega}, \frac{A}{\omega}, B \right) = \frac{\sigma^2}{\omega^2} \tilde{\lambda}_{\text{ave}} \left(\frac{\omega^2 \beta}{\sigma^2}, 1, 1, \frac{\omega B}{A} \right),$$

valid for $\omega^2\beta/\sigma^2 > -1/4$. Therefore in our numerical computations in Section 7 it suffices to take $\omega = \sigma = 1$ and $\beta > -1/4$. We note that in Eq. (4) A has dimension $(\text{length})^{-2}(\text{time})^{-2}$ and B has dimension $(\text{length})^{-2}(\text{time})^{-1}$ and ω has dimension $(\text{time})^{-1}$ so that $\omega B/A$ is a dimensionless quantity. It measures the size of the nonlinear dissipation term $Bx^2\dot{x}$ relative to the size of the nonlinear restoring force Ax^3.

Now consider the bifurcation scenario for Eq. (6) where the coefficients are varied in such a way that $\lambda \searrow 0$. Suppose for example that σ and ω are fixed and $\beta \searrow -\sigma^2/4\omega^2$. By Baxendale [8] Theorem 5.1 the limit

$$\Gamma\left(\frac{\sigma}{\omega}, \frac{A}{\omega}, B\right) \equiv \lim_{\beta \searrow -\sigma^2/4\omega^2} \frac{\tilde{\lambda}_{\text{ave}}(\beta, \frac{\sigma}{\omega}, \frac{A}{\omega}, B)}{\lambda_{\text{ave}}(\beta, \frac{\sigma}{\omega})}$$

exists. By the results above we get

$$\Gamma\left(\frac{\sigma}{\omega}, \frac{A}{\omega}, B\right) = \lim_{\beta \searrow -\sigma^2/4\omega^2} \frac{\tilde{\lambda}_{\text{ave}}(\frac{\omega^2\beta}{\sigma^2}, 1, 1, \frac{\omega B}{A})}{\frac{\omega^2\beta}{2\sigma^2} + \frac{1}{8}}$$

$$= \lim_{\beta \searrow -1/4} \frac{\tilde{\lambda}_{\text{ave}}(\beta, 1, 1, \frac{\omega B}{A})}{\frac{\beta}{2} + \frac{1}{8}} = \Gamma\left(1, 1, \frac{\omega B}{A}\right)$$

so that the value of Γ at the bifurcation point depends only on the ratio $\omega B/A$. Clearly, if $\Gamma > 0$ then we have $\tilde{\lambda} > 0$ for small positive λ, while if $\Gamma < 0$ we have $\tilde{\lambda} < 0$ for small positive λ.

7. Numerical results

Define $\chi \in (0, \pi/4)$ so that $\tan 2\chi = 3A/\omega B$. Write $s = 1/r^2$ and $\bar{\psi} = \psi + \chi$. Then Eqs. (41) and (42) give

$$\mathrm{d}s_t = \left(-\beta s_t + \frac{B}{4}\right)\mathrm{d}t - \frac{\sigma}{\sqrt{2}}s_t\,\mathrm{d}B_t^1 \tag{45}$$

and

$$\mathrm{d}\bar{\psi}_t = \frac{1}{8\omega s_t}\left(3A + \sqrt{9A^2 + \omega^2B^2}\sin 2\bar{\psi}_t\right)\mathrm{d}t$$
$$- \frac{\sigma}{2\sqrt{2}\omega}\sin 2(\bar{\psi}_t - \chi)\,\mathrm{d}B_t^1 + \frac{\sigma}{2\sqrt{2}\omega}(-1 + \cos 2(\bar{\psi}_t - \chi))\,\mathrm{d}B_t^2. \tag{46}$$

Also

$$\bar{Q}(r,\psi) = -\frac{\sqrt{9A^2 + \omega^2B^2}}{8\omega}\frac{\cos 2\bar{\psi}}{s}.$$

The following numerical results use a first order Euler scheme to solve system (45)–(46). The discrete time average of $(\cos 2\bar{\psi}_t)/s_t$ is then used to give an estimate for the integral of $\bar{Q}$ with respect to the invariant measure $\bar{\nu}$.

The computations used values $\omega = 1$ and $\sigma = 1$. The parameter β was varied from -0.245 to -0.005 in steps of 0.005. In Figs. 1 and 2 the dotted line shows the exact

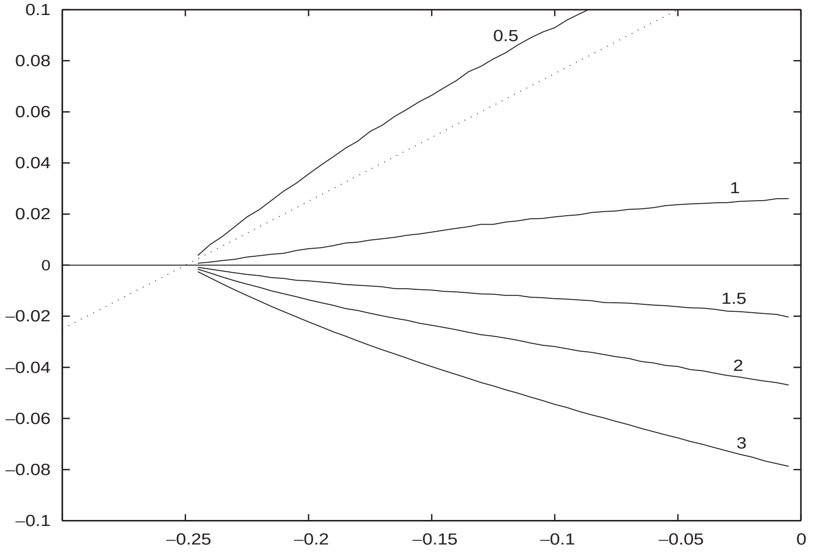

Fig. 1. Lyapunov exponent as a function of β when $\sigma^2/\omega^2 = 1$. Dotted line: $\lambda = \beta/2 + 1/8$. Solid lines: $\tilde{\lambda}$ for $\omega B/A = 0.5$, 1, 1.5, 2 and 3.

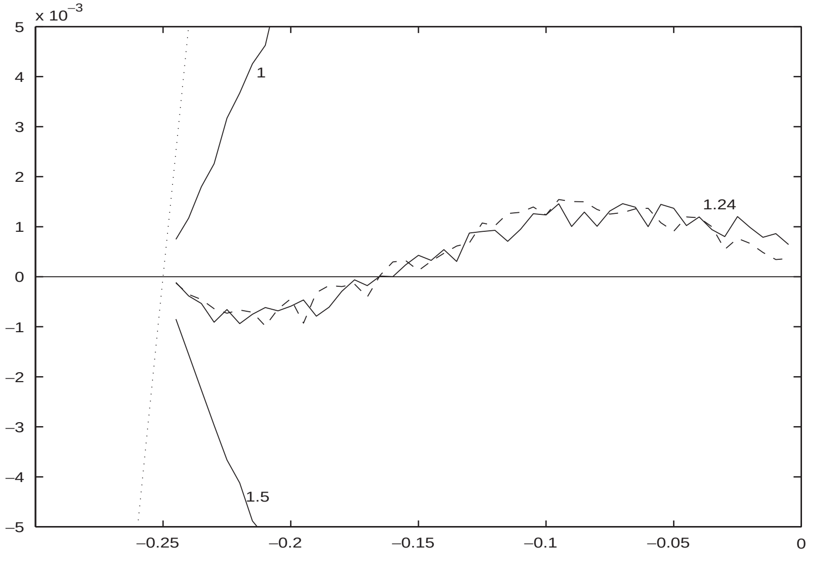

Fig. 2. Lyapunov exponent as a function of β when $\sigma^2/\omega^2 = 1$. Dotted line: $\lambda = \beta/2 + 1/8$. Solid lines: $\tilde{\lambda}$ for $\omega B/A = 1$, 1.24 and 1.5 with $h = 0.001$. Dashed line: $\tilde{\lambda}$ for $\omega B/A = 1.24$ with $h = 0.0005$.

value $\lambda = \beta/2 + 1/8$. The solid lines show the results of simulation of $\tilde{\lambda}$ with time step $h = 0.001$ and number of iterations $M = 7.8 \times 10^8$ (so that total time $= Mh = 7.8 \times 10^5$). The dashed line shows the results of simulation of $\tilde{\lambda}$ with time step $h = 0.0005$ and number of iterations $M = 15.6 \times 10^8$. Fig. 1 shows $\tilde{\lambda}$ for the values $\omega B/A = 0.5$, 1, 1.5, 2 and 3. Fig. 2 shows $\tilde{\lambda}$ for the values $\omega B/A = 1$, 1.24 and 1.5.

We see from the numerical data in Fig. 1 that $\tilde{\lambda} > 0$ (instability along trajectories for (6)) when $\omega B/A \leqslant 1$ and $-1/4 < \beta < 0$. Also we have $\tilde{\lambda} < 0$ (stability along trajectories for (6)) when $\omega B/A \geqslant 1.5$ and $-1/4 < \beta < 0$. Notice that the data is consistent with the existence of the limits $\Gamma(1, 1, \omega B/A)$ as $\beta \searrow -0.25$.

Finally consider the numerical data in Fig. 2 for $\omega B/A = 1.24$. (Notice the expanded vertical scale in Fig. 2.) The apparent randomness in the data is caused by the finiteness of the number of iterations M. But even in the theoretical limit as $M \to \infty$ there is an error in our calculations because of the finite time step h in the Euler scheme. Here we are dealing with values of $\tilde{\lambda}$ which are of the same order as the time step h, and a theoretical analysis (see Talay [28]) shows that the possible error in the computed value of $\tilde{\lambda}$ is of order h. Thus we cannot be sure of the sign of the true value of $\tilde{\lambda}$ when $\omega B/A = 1.24$. However, Fig. 2 shows that halving the time step (from 0.001 to 0.0005) has little effect on the data. Thus it is reasonable to believe that the graph of the true value of $\tilde{\lambda}$ will be close to a smoothed out version of the solid (or dashed) line. Therefore our calculations strongly suggest the following:

Conjecture. There exists a value β_c, approximately equal to 0.17 such that $\tilde{\lambda}_{\text{ave}}(\beta, 1, 1, 1.24) < 0$ for $-1/4 < \beta < -\beta_0$ and $\tilde{\lambda}_{\text{ave}}(\beta, 1, 1, 1.24) > 0$ for $-\beta_0 < \beta < 0$. Now consider the change in behavior of system (6) with $\omega B/A = 1.24$ as β is increased from $-\infty$ to 0. (Recall that $-\beta$ represents friction, so we are decreasing the dissipation due to friction.) For $\beta < -\sigma^2/4\omega^2$ the fixed point 0 is almost surely stable. For $-\sigma^2/4\omega^2 < \beta < \beta_0\sigma^2/\omega^2$ the fixed point 0 is almost surely stable, but the system is positive recurrent on $\mathbf{R}^2/\{0\}$ and has stability along trajectories. The random invariant measure for the flow on $\mathbf{R}^2/\{0\}$ is a random Dirac measure. For $-\beta_c\sigma^2/\omega^2 < \beta < 0$ the system loses stability along its trajectories, and the random invariant measure now becomes non-atomic and has Hausdorff dimension greater than 1. Thus we observe two qualitative changes in the dynamic behavior (or "dynamic bifurcations") for system (6), the first at $\beta = -\sigma^2/4\omega^2$ and the other at $\beta = -\beta_0\sigma^2/\omega^2$.

Acknowledgements

I would like to thank Levon Goukasian (University of Southern California) for his assistance with the numerical computations in Section 7.

Appendix A. Stochastic averaging

In this appendix we extend some stochastic averaging results of Khasminskii [15]. The main novelty is the replacement of global boundedness assumptions on the coefficients by assumptions of the form of B2 and B3 below. Since our main concern is Theorem A.3 which deals with the convergence of invariant measures, we follow the methods of Khasminskii, concentrating on the Markov semigroups, instead of using the powerful martingale methods of Papanicolaou et al. [23].

Consider for $\varepsilon > 0$ the process $(\varphi_t^\varepsilon, x_t^\varepsilon)$ with values in $\mathbf{S}^1 \times \mathbf{R}^d$ given by the equations

$$\begin{aligned} d\varphi_t^\varepsilon &= \omega\, dt + \varepsilon^2 b^0(\varphi_t^\varepsilon, x_t^\varepsilon)\, dt + \varepsilon \sum_{\alpha=1}^k \sigma_\alpha^0(\varphi_t^\varepsilon, x_t^\varepsilon)\, dW_t^\alpha, \\ dx_t^{\varepsilon,i} &= \varepsilon^2 b^i(\varphi_t^\varepsilon, x_t^\varepsilon)\, dt + \varepsilon \sum_{\alpha=1}^k \sigma_\alpha^i(\varphi_t^\varepsilon, x_t^\varepsilon)\, dW_t^\alpha \quad \text{for } 1 \leqslant i \leqslant d. \end{aligned} \tag{A.1}$$

Write $a^{ij}(\varphi, x) = \sum_{\alpha=1}^k \sigma_\alpha^i(\varphi, x)\sigma_\alpha^j(\varphi, x)$. With the convention that $\partial/\partial\varphi = \partial/\partial x^0$, let L denote the operator

$$LF(\varphi, x) = \sum_{i=0}^d b^i(\varphi, x)\frac{\partial F}{\partial x^i}(\varphi, x) + \frac{1}{2}\sum_{i,j=0}^d a^{ij}(\varphi, x)\frac{\partial^2 F}{\partial x^i \partial x^j}(\varphi, x).$$

Then $(\varphi_t^\varepsilon, x_t^\varepsilon)$ has generator $L_\varepsilon = \omega\partial/\partial\varphi + \varepsilon^2 L$. Denote by $\tilde{L}$ the operator obtained from L by averaging all of its coefficients over φ. Thus

$$\tilde{L}F(\varphi, x) = \sum_{i=0}^d \bar{b}^i(x)\frac{\partial F}{\partial x^i}(\varphi, x) + \frac{1}{2}\sum_{i,j=0}^d \bar{a}^{ij}(x)\frac{\partial^2 F}{\partial x^i \partial x^j}(\varphi, x),$$

where

$$\bar{b}^i(x) = \frac{1}{2\pi}\int_0^{2\pi} b^i(\varphi, x)\, d\varphi$$

and

$$\bar{a}^{ij}(x) = \frac{1}{2\pi}\int_0^{2\pi} a^{ij}(\varphi, x)\, d\varphi.$$

Restricting to the x variable only we get the differential operator $\bar{L}$, say, given by

$$\bar{L}F(x) = \sum_{i=1}^d \bar{b}^i(x)\frac{\partial F}{\partial x^i}(x) + \frac{1}{2}\sum_{i,j=1}^d \bar{a}^{ij}(x)\frac{\partial^2 F}{\partial x^i \partial x^j}(x).$$

At this point we list the assumptions:

B1. The functions b^i and σ_α^i, $0 \leqslant i \leqslant d$, $1 \leqslant \alpha \leqslant k$, are C^2 functions of (φ, x).

B2. There exists $c < \infty$ and $\varepsilon_0 > 0$ and C^2 functions $V_\varepsilon(\varphi, x)$ for each $0 < \varepsilon \leqslant \varepsilon_0$ such that $\inf_\varepsilon \inf_\varphi V_\varepsilon(\varphi, x) \to \infty$ as $\|x\| \to \infty$ and $\sup_\varepsilon \sup_{\varphi, x \in K} V_\varepsilon(\varphi, x) < \infty$ for each compact $K \subset \mathbf{R}^d$ and

$$L_\varepsilon V_\varepsilon(\varphi, x) \leqslant \varepsilon^2 c V_\varepsilon(\varphi, x) \quad \text{for all } \varepsilon > 0$$

B3. There exists $c < \infty$ and a C^2 function $V_0(x)$ such that $V_0(x) \to \infty$ as $\|x\| \to \infty$ and

$$\bar{L}V_0(x) \leqslant cV_0(x).$$

The assumption B1 implies that the process $(\varphi_t^\varepsilon, x_t^\varepsilon)$ given by (A.1) has a solution up to some explosion time, and then the assumption B2 implies that the explosion time is almost surely infinite, see Khasminskii [16, Theorem III.4.1]. We shall use P_t^ε to denote the Markov semigroup associated with (A.1). Notice that B1 also implies that $\tilde{L}$ has C^2 coefficients. This together with B3 gives the existence of a non-explosive Markov semigroup (Φ_t, X_t), say, with generator $\tilde{L}$ and associated Markov semigroup $\tilde{P}_t$.

The following Theorem and Corollary are extensions to a wider class of coefficients of Theorem 2 and the discussion in Section 4.4 of Khasminskii [15].

Theorem A.1. *Assume* B1–B3. *For any bounded continuous $F(\varphi,x)$ and $T > 0$ and any compact $K \subset \mathbf{R}$ we have*

$$\lim_{\varepsilon\to 0} \sup_{t\leqslant T} \sup_{\varphi,\ x\in K} |P_{t/\varepsilon^2}^\varepsilon F(\varphi,x) - \tilde{P}_t F(\varphi + \omega t/\varepsilon^2, x)| = 0. \tag{A.2}$$

Proof. First we will prove the result under the extra assumptions that F is C^2 and that there exists a compact $K_1 \subset \mathbf{R}^n$ such that b^i and σ_α^i and F all have compact support inside $\mathbf{S}^1 \times K_1$. For $0 \leqslant s \leqslant T/\varepsilon^2$ define

$$u_\varepsilon(\varphi,x,s) = E[F(\varphi_{T/\varepsilon^2}^\varepsilon, x_{T/\varepsilon^2}^\varepsilon) | \varphi_s^\varepsilon = \varphi,\ x_s^\varepsilon = x].$$

Then $\lim_{s\to T/\varepsilon^2} u_\varepsilon(\varphi,x,s) = F(\varphi,x)$ and

$$\frac{\partial u_\varepsilon}{\partial s} + \omega\frac{\partial u_\varepsilon}{\partial \varphi} + \varepsilon^2 L u_\varepsilon = 0.$$

Write $v_\varepsilon(\varphi,x,s) = u_\varepsilon(\varphi + \omega(s - T/\varepsilon^2), x, s)$. Then $\lim_{s\to T/\varepsilon^2} v_\varepsilon(\varphi,x,s) = F(\varphi,x)$ and

$$\frac{\partial v_\varepsilon}{\partial s}(\varphi,x,s) + \varepsilon^2 \sum_{i=0}^d b^i(\varphi + \omega(s - T/\varepsilon^2), x)\frac{\partial v_\varepsilon}{\partial x_i}(\varphi,x,s)$$

$$+ \frac{\varepsilon^2}{2}\sum_{i,j=0}^d a^{ij}(\varphi + \omega(s - T/\varepsilon^2), x)\frac{\partial^2 v_\varepsilon}{\partial x^i \partial x^j}(\varphi,x,s) = 0.$$

For $0 \leqslant t \leqslant T$ define $v(\varphi,x,t) = E[F(\Phi_T, X_T) | \Phi_t = \varphi, X_t = x]$. Then $\lim_{t\to T} v(\varphi,x,t) = F(\varphi,x)$ and

$$\frac{\partial v}{\partial t}(\varphi,x,t) + \tilde{L}v(\varphi,x,t) = 0.$$

The functions u_ε and v_ε and v are all zero when $x \notin K_1$. Also, by a result of Oleinik [22] (see also Stroock and Varadhan [27, Appendix]), the function $(\varphi,x) \mapsto v(\varphi,x,t)$ is C^2 for each $t \in [0,T]$ with bounded second derivatives. The method of proof of [15, Theorem 2] now gives

$$\lim_{\varepsilon\to 0}\sup_{t\in[0,T]}\sup_{\varphi,\ x}|v_\varepsilon(\varphi,x,t/\varepsilon^2)-v(\varphi,x,t)|=0.$$

Recalling the definitions of v_ε and v, and replacing t with $T-t$, we get

$$\lim_{\varepsilon\to 0}\sup_{t\leqslant T}\sup_{\varphi,x}|P^\varepsilon_{t/\varepsilon^2}F(\varphi-\omega t/\varepsilon^2,x)-\tilde{P}_tF(\varphi,x)|=0.$$

Result (A.2) follows by replacing φ with $\varphi+\omega t/\varepsilon^2$.

Now we remove the extra assumptions that F is C^2 and that b^i and σ^i_α and F have compact support. Let G_n, $n\geqslant 1$, be compact sets with $G_n\subset \mathrm{int}(G_{n+1})$ and $\cup_{n\geqslant 1}G_n=\mathbf{R}^d$. Then $a_n\equiv \inf_\varepsilon\inf_{\varphi,y\notin G_n}V_\varepsilon(\varphi,y)\nearrow\infty$ as $n\to\infty$. For the process $(\varphi^\varepsilon_t,x^\varepsilon_t)$ given by (A.1) let $\tau^\varepsilon_n=\inf\{t\geqslant 0: x^\varepsilon_t\notin G_n\}$. The condition $L_\varepsilon V_e\leqslant\varepsilon^2cV_\varepsilon$ implies that $\mathrm{e}^{-\varepsilon^2c(t\wedge\tau^\varepsilon_n)}V_\varepsilon(x^\varepsilon_{(t\wedge\tau^\varepsilon_n)})$ is a supermartingale, and this in turn implies $P^{(\varphi,x)}(\tau^\varepsilon_n\leqslant t)\leqslant \mathrm{e}^{-\varepsilon^2ct}V_\varepsilon(\varphi,x)/a_n$. It follows that

$$\sup_{\varepsilon>0}\sup_{t\leqslant T}\sup_{\varphi,x\in K}P^{(\varphi,x)}(\tau^\varepsilon_n\leqslant t/\varepsilon^2)\to 0\quad\text{as } n\to\infty \tag{A.3}$$

for each $T>0$ and compact $K\subset\mathbf{R}^d$. Similarly, for the process (Φ_t,X_t) let $\sigma_n=\inf\{t\geqslant 0: X_t\notin G_n\}$. The estimate $\bar{L}V_0(x)\leqslant cV_0(x)$ implies that

$$\sup_{t\leqslant T}\sup_{\varphi,x\in K}P^{(\varphi,x)}(\sigma_n\leqslant t)\to 0\quad\text{as } n\to\infty \tag{A.4}$$

for each $T>0$ and compact $K\subset\mathbf{R}^d$.

For any $n\geqslant 1$ let $b^i_n(\varphi,x)$ and $\sigma^i_{\alpha,n}(\varphi,x)$ be compactly supported functions which agree with $b^i(\varphi,x)$ and $\sigma^i_\alpha(\varphi,x)$ on the set $\mathbf{S}^1\times G_n$. Let $P^\varepsilon_{t,n}$ and $\tilde{P}_{t,n}$ be the Markov semigroups determined by the functions b^i_n and $\sigma^i_{\alpha,n}$ in the same way that P^ε_t and $\tilde{P}_t$ were determined by the functions b^i and σ^i_α. Moreover, there exists $c_1<\infty$ and for each $n\geqslant 1$ there exists a compactly supported C^2 function F_n with the properties that $\sup_{\varphi,x\in G_n}|F_n(\varphi,x)-F(\varphi,x)|\leqslant 1/n$ and $\sup_{\varphi,x}|F_n(\varphi,x)-F(\varphi,x)|\leqslant c_1$. Then for any compact $K\subset G_n$ we have

$$\sup_{t\leqslant T}\sup_{\varphi,x\in K}|P^\varepsilon_{t/\varepsilon^2,n}F_n(\varphi,t)-P^\varepsilon_{t/\varepsilon^2}F(\varphi,t)|\leqslant\frac{1}{n}+c_1\sup_{t\leqslant T}\sup_{\varphi,x\in K}P^{(\varphi,x)}(\tau^\varepsilon_n\leqslant t/\varepsilon^2). \tag{A.5}$$

Similarly

$$\sup_{t\leqslant T}\sup_{\varphi,x\in K}|\tilde{P}_{t,n}F_n(\varphi+\omega t/\varepsilon^2,x)-\tilde{P}_tF(\varphi+\omega t/\varepsilon^2,x)|$$

$$\leqslant\frac{1}{n}+c_1\sup_{t\leqslant T}\sup_{\varphi,x\in K}P^{(\varphi,x)}(\sigma_n\leqslant t). \tag{A.6}$$

The estimates (A.3) and (A.4) imply that the right sides of (A.5) and (A.6) tend to 0 as $n\to\infty$ uniformly in ε. Our earlier calculations showed that

$$\lim_{\varepsilon\to 0}\sup_{t\leqslant T}\sup_{\varphi,x}|P^\varepsilon_{t/\varepsilon^2,n}F_n(\varphi,x)-\tilde{P}_{t,n}F_n(\varphi+\omega t/\varepsilon^2,x)|=0$$

for each $n\geqslant 1$. Putting together the last three estimates gives the required result. □

Corollary A.1. *Assume* B1–B3. *For any continuous bounded $F(\varphi,x)$ and $0 \leqslant T_1 < T_2$ and any compact set $K \subset \mathbf{R}^d$ we have*

$$\lim_{\varepsilon\to 0} \sup_{\varphi,x\in K} \left| \int_{T_1}^{T_2} P^{\varepsilon}_{t/\varepsilon^2} F(\varphi,x)\,\mathrm{d}t - \int_{T_1}^{T_2} \left(\frac{1}{2\pi} \int_0^{2\pi} \tilde{P}_t F(\psi,x)\,\mathrm{d}\psi \right) \mathrm{d}t \right| = 0.$$

Proof. By Theorem A.1, it suffices to show that

$$\lim_{\varepsilon\to 0} \sup_{\varphi,x\in K} \left| \int_{T_1}^{T_2} \tilde{P}_t F(\varphi + \omega t/\varepsilon^2, x)\,\mathrm{d}t - \int_{T_1}^{T_2} \left(\frac{1}{2\pi} \int_0^{2\pi} \tilde{P}_t F(\psi,x)\,\mathrm{d}\psi \right) \mathrm{d}t \right| = 0,$$

and this estimate is an elementary consequence of the fact that $\tilde{P}_t F(\varphi,x)$ is a continuous function of t and (φ,x). $\square$

Since all the coefficients of $\tilde{L}$ depend only on x, it follows that X_t is itself a Markov process with generator $\bar{L}$ and corresponding semigroup $\bar{P}_t$, say. Also, the joint law of $\Phi_t - \Phi_0$ and X_t is independent of Φ_0. It follows that for any bounded continuous $F(\varphi,x)$ we have

$$\int \tilde{P}_t F(\varphi,x)\,\mathrm{d}m(\varphi) = \bar{P}_t \bar{F}(x),$$

where m denotes the uniform probability measure (Haar measure) on $\mathbf{S}^1$ and $\bar{F}(x) = \int F(\varphi,x)\,\mathrm{d}m(\varphi)$.

The following result gives the more usual statement of stochastic averaging in terms of the convergence in law of certain processes. In this result we consider only the slow variable x^{ε}_t.

Theorem A.2. *Assume* B1–B3. *For any $\varphi \in \mathbf{S}^1$ and $x \in \mathbf{R}^d$ and $T < \infty$ the process $\{x_{t/\varepsilon^2} : 0 \leqslant t \leqslant T\}$ given by* (A.1) *with $\varphi_0 = \varphi$, $x_0 = x$ converges weakly in $C([0,T],\mathbf{R}^d)$ to the process $\{X_t : 0 \leqslant t \leqslant T\}$ with generator $\bar{L}$ started at $X_0 = x$.*

Proof. This is an application of Ethier and Kurtz [12, Chapter 4, Corollary 8.7]. The condition (8.35) of [12] is guaranteed by (A.3), and the condition (8.37) of [12] is exactly Theorem A.1 with F a function of x only. Notice finally that for continuous processes, weak convergence in the topology of $D([0,T],\mathbf{R}^d)$ implies weak convergence in the topology of $C([0,T],\mathbf{R}^d)$. $\square$

The result above is a statement about convergence on fixed finite time intervals and cannot be used directly to make assertions about limiting behavior as time goes to infinity. The following result, based on Section 5 of Khasminskii [15], deals with convergence of invariant probability measures.

Theorem A.3. *Assume* B1–B3. *Assume also*

(i) *the process $(\varphi^{\varepsilon}_t, x^{\varepsilon}_t)$ given by* (A.1) *has an invariant probability measure $\tilde{\mu}_{\varepsilon}$ and the measures $\tilde{\mu}_{\varepsilon}$, $0 < \varepsilon \leqslant \varepsilon_0$, are tight for some $\varepsilon_0 > 0$; and*

(ii) *the process X_t with generator $\bar{L}$ has at most one invariant probability measure.*

Then the process X_t with generator $\bar{L}$ has a unique invariant probability measure μ, say, and the probability measures $\tilde{\mu}_\varepsilon$ on $\mathbf{S}^1 \times \mathbf{R}^d$ converge weakly to the product measure $m \times \mu$, where m denotes Haar measure on $\mathbf{S}^1$.

Proof. Suppose that $\tilde{\mu}_{\varepsilon_i}$ converges weakly to a probability measure $\tilde{\mu}$ for some sequence $\varepsilon_i \to 0$. For any bounded continuous function F on $\mathbf{S}^1 \times \mathbf{R}^d$, let $\bar{F}(x) = \int F(\varphi, x)\,\mathrm{d}m(\varphi)$. The functions $P^\varepsilon_{t/\varepsilon^2} F(\varphi, x)$ and $\bar{P}_t \bar{F}(x)$ are bounded continuous functions of t and (φ, x). Therefore for any $T > 0$ the functions $(1/T) \int_0^T P^\varepsilon_{t/\varepsilon^2} F(\varphi, x)\,\mathrm{d}t$ and $(1/T) \int_0^T \bar{P}_t \bar{F}(x)\,\mathrm{d}t$ are bounded continuous functions of (φ, x). Corollary A.1 implies that

$$\frac{1}{T} \int_0^T P^\varepsilon_{t/\varepsilon^2} F(\varphi, x)\,\mathrm{d}t \to \frac{1}{T} \int_0^T \left(\int \bar{P}_t F(\psi\ , x)\,\mathrm{d}m(\psi) \right) \mathrm{d}t = \frac{1}{T} \int_0^T \bar{P}_t \bar{F}(x)\,\mathrm{d}t$$

uniformly on compact subsets of $\mathbf{S}^1 \times \mathbf{R}^d$ as $\varepsilon \to 0$. The weak convergence $\tilde{\mu}_{\varepsilon_i} \to \tilde{\mu}$ then gives

$$\lim_i \int \left(\frac{1}{T} \int_0^T P^{\varepsilon_i}_{t/\varepsilon_i^2} F(\varphi, x)\,\mathrm{d}t \right) \mathrm{d}\tilde{\mu}_{\varepsilon_i}(\varphi, x) = \int \left(\frac{1}{T} \int_0^T \bar{P}_t \bar{F}(x)\,\mathrm{d}t \right) \mathrm{d}\tilde{\mu}(\varphi, x).$$

Then

$$\begin{aligned} \int F(\varphi, x)\,\mathrm{d}\tilde{\mu}(\varphi, x) &= \lim_i \int F(\varphi, x)\,\mathrm{d}\tilde{\mu}_{\varepsilon_i}(\varphi, x) \\ &= \lim_i \int \left(\frac{1}{T} \int_0^T P^{\varepsilon_i}_{t/\varepsilon_i^2} F(\varphi, x)\,\mathrm{d}t \right) \mathrm{d}\tilde{\mu}_{\varepsilon_i}(\varphi, x) \\ &= \int \left(\frac{1}{T} \int_0^T \bar{P}_t \bar{F}(x)\,\mathrm{d}t \right) \mathrm{d}\tilde{\mu}(\varphi, x). \end{aligned}$$

Since replacing $F(\varphi, x)$ by $\bar{F}(x)$ on the left side above does not change the right side, it follows that $\tilde{\mu} = m \times \mu$ for some probability measure μ on $\mathbf{R}^d$. Then the equation above gives

$$\int f(x)\,\mathrm{d}\mu(x) = \int \left(\frac{1}{T} \int_0^T \bar{P}_t f(x)\,\mathrm{d}t \right) \mathrm{d}\mu(x)$$

for all bounded continuous functions f on $\mathbf{R}^d$. It follows that μ must be invariant for the X_t process. By assumption μ is unique, and so the limit probability measure $\tilde{\mu} = m \times \mu$ is unique. The result now follows by the tightness of the $\tilde{\mu}_\varepsilon$. $\square$

Remark. All the results in this section remain true if the space $\mathbf{R}^d$ is replaced by the space $\mathbf{R}^{d_1} \times \mathbf{T}^{d_2}$, where $\mathbf{T}^d$ denotes the d-dimensional torus.

References

[1] L. Arnold, Random Dynamical Systems, Springer, Berlin, 1998.

[2] L. Arnold, R. Khasminskii, Stability index for nonlinear stochastic differential equations, Proc. Symp. Pure Math. 57 (1995) 543–551.
[3] L. Arnold, W. Kliemann, On unique ergodicity for degenerate diffusions, Stochastics 21 (1987) 41–61.
[4] L. Arnold, L. San Martin, A control problem related to the Lyapunov spectrum of stochastic flows, Mat. Apl. Comput. 5 (1986) 31–64.
[5] L. Arnold, N. Sri Namachchivaya, K. Schenk-Hoppé, Toward an understanding of the stochastic Hopf bifurcation: a case study, Internat. J. Bifur. Chaos 6 (1996) 1947–1975.
[6] E. Auslender, G. Mil'shtein, Asymptotic expansions of the Liapunov index for linear stochastic systems with small noise, J. Appl. Math. Mech. 46 (1982) 277–283.
[7] P.H. Baxendale, Invariant measures for nonlinear stochastic differential equations, in: L. Arnold, H. Crauel, J.-P. Eckmann (Eds.), Lyapunov Exponents. Proc. Oberwolfach 1990, Lecture Notes in Mathematics, Vol. 1486, Springer, Berlin, 1991, pp. 123–140.
[8] P.H. Baxendale, Stability along trajectories at a stochastic bifurcation point, in: H. Crauel, M. Gundlach (Eds.), Stochastic Dynamics, Springer, Berlin, 1999, pp. 1–25.
[9] P.H. Baxendale, L. Goukasian, Lyapunov exponents for small random perturbations of Hamiltonian systems, Ann. Probab. 30 (2002) 101–134.
[10] P.H. Baxendale, D.W. Stroock, Large deviations and stochastic flows of diffeomorphisms, Probab. Theory Related Fields 80 (1988) 169–215.
[11] A.P. Carverhill, A formula for the Lyapunov numbers of a stochastic flow. Application to a perturbation theorem, Stochastics 14 (1985) 209–226.
[12] S. Ethier, T. Kurtz, Markov Processes, Wiley, New York, 1986.
[13] P. Imkeller, C. Lederer, An explicit description of the Lyapunov exponents of the noisy damped harmonic oscillator, Dynamics Stability Systems 14 (1999) 385–405.
[14] H. Keller, G. Ochs, Numerical approximation of random attractors, in: H. Crauel, M. Gundlach (Eds.), Stochastic dynamics, Springer, Berlin, 1999, pp. 93–115.
[15] R.Z. Khasminskii, Principle of averaging for parabolic and elliptic differential equations and for Markov processes with small diffusion, Theory Probab. Appl. 8 (1963) 1–21.
[16] R.Z. Khasminskii, Necessary and sufficient conditions for the asymptotic stability of linear stochastic systems, Theory Probab. Appl. 12 (1967) 144–147.
[17] R.Z. Khasminskii, Stochastic Stability of Differential Equations, Sijthoff and Noordhoff, Alphen aan den Rijn, 1980.
[18] R. Khasminskii, N. Moshchuk, Moment Lyapunov exponent and stability index for linear conservative system with small random perturbation, SIAM J. Appl. Math. 58 (1998) 245–256.
[19] F. Kozin, S. Prodromou, Necessary and sufficient conditions for almost sure sample stability of linear Itô equations, SIAM J. Appl. Math. 21 (1971) 413–424.
[20] H. Kunita, Stochastic Flows and Stochastic Differential Equations, Cambridge University Press, Cambridge, 1990.
[21] S. Meyn, R. Tweedie, Stability of Markovian processes III: Foster-Lyapunov criteria for continuous time processes, Adv. in Appl. Prob. 25 (1993) 518–548.
[22] O.A. Oleinik, Alcuni risultati sulle equazioni lineari e quasi lineari ellittico-paraboliche a derivate parziali del secondo ordine, Atti Accad. Naz. Lincei Rend. Cl. Sci. Fis. Mat. Natur. (8) 40 (1966) 775–784.
[23] G.G. Papanicolaou, D.W. Stroock, S.R.S. Varadhan, Martingale approach to some limit theorems, 1976 Duke Turbulence Conference, Duke Univ. Math. Ser. III (1997) (Chapter VI) 1–120.
[24] E. Pardoux, V. Wihstutz, Lyapunov exponent and rotation number of two-dimensional linear stochastic systems with small diffusion, SIAM J. Appl. Math. 48 (1988) 442–457.
[25] D. Ruelle, Ergodic theory of differentiable dynamical systems, Publ. Math. IHES 50 (1979) 275–305.
[26] K. Schenk-Hoppé, Bifurcation scenarios of the noisy Duffing–van der Pol oscillator, Nonlinear Dynamics 11 (1996) 255–274.
[27] D.W. Stroock, S.R.S. Varadhan, On degenerate elliptic-parabolic operators of second order and their associated diffusions, Comm. Pure Appl. Math. 25 (1972) 651–713.
[28] D. Talay, The Lyapunov exponent for the Euler scheme for stochastic differential equations, in: H. Crauel, M. Gundlach (Eds.), Stochastic Dynamics, Springer, Berlin, 1999, pp. 241–258.